

P

Access DB# 27616

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

FOR OFFICIAL USE ONLY

STAFF USE ONLY		Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____	
Searcher Phone #: _____	AA Sequence (#) <u>1</u>	Dialog _____	
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____	
Date Searcher Picked Up: <u>11/16/02</u>	Bibliographic _____	Dr.Link _____	
Date Completed: <u>11/17/02</u>	Litigation _____	Lexis/Nexis _____	
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>OK</u>	
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____	
Online Time: _____	Other _____	Other (specify) _____	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 16, 2000, 15:36:10 ; Search time 15.12 Seconds

(without alignments)
13.569 Million cell updates/sec

Title: US-09-308-237b-1

Sequence: 35

Scoring table: 1 PROQ6 6

Scoring table: BLOSUM62

Searched: 268485 seqs, 3419375 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq.36.*
1: /SIDSL/gcgdata/geneSeq/geneSeq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneSeq/geneSeq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneSeq/geneSeq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneSeq/geneSeq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneSeq/geneSeq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneSeq/geneSeq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneSeq/geneSeq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneSeq/geneSeq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneSeq/geneSeq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneSeq/geneSeq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneSeq/geneSeq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneSeq/geneSeq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneSeq/geneSeq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneSeq/geneSeq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneSeq/geneSeq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneSeq/geneSeq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneSeq/geneSeq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneSeq/geneSeq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneSeq/geneSeq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneSeq/geneSeq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneSeq/geneSeq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	573	16	R80633
2	32	91.4	797	20	Y36955
3	31	88.6	300	21	Y4616
4	31	88.6	350	18	W14532
5	31	88.6	374	15	R45339
6	31	88.6	374	18	W13542
7	30	85.7	20	19	Y20992
8	29	82.9	13	13	R28833
9	29	82.9	41	19	W42445
10	29	82.9	74	6	P50761
11	29	82.9	125	19	Y21354
12	29	82.9	127	20	Y73845

13 82.9 135 21 Y4649
14 82.9 162 6 P51201
15 29 82.9 262 21 Y91338
16 29 82.9 326 19 W36985
17 29 82.9 369 19 Y70224
18 29 82.9 459 19 W56738
19 29 82.9 460 21 Y94991
20 29 82.9 485 19 W5462
21 29 82.9 485 20 Y27283
22 29 82.9 567 20 W78489
23 29 82.9 747 19 W60487
24 29 82.9 755 20 W81365
25 29 82.9 763 19 W60591
26 29 82.9 1078 13 R28963
27 29 82.9 1093 14 R41001
28 29 82.9 1135 21 Y94923
29 82.9 1661 19 W76041
30 29 82.9 2414 16 R84882
31 29 82.9 2414 19 W40057
32 29 82.9 2414 19 Y30717
33 28 80.0 63 20 Y48601
34 28 80.0 144 20 Y28974
35 28 80.0 344 18 R83797
36 28 80.0 400 8 P70654
37 28 80.0 405 21 Y70440
38 28 80.0 531 21 Y70439
39 28 80.0 547 19 W40112
40 28 80.0 581 20 Y34560
41 28 80.0 601 20 Y34423
42 28 80.0 620 21 Y70438
43 28 80.0 622 21 Y70437
44 28 80.0 664 12 R14666
45 28 80.0 795 19 W72062

ALIGNMENTS

RESULT 1

ID R80633 standard; Protein: 573 AA.

XX AC R80633;

XX DT 27-FEB-1996 (first entry)

XX DE Clone 3-2 encoded protein.

XX KW Anti-Mullerian hormone receptor; anti-Mullerian hormone; AMH;

XX KW receptor; antibody; therapy; H1; 2B10; tumour;

XX KW transforming growth factor-beta; TGF-beta; bone morphogenic protein;

XX KW activin; Mullerian inhibiting substance; MIS.

XX OS Homo sapiens.

XX FH Key

XX FT Peptide

XX FT Protein

XX FT Domain

XX FT Misc-difference

XX FT Misc-difference

XX FT Region

XX FT Domain

XX FT Misc-difference

XX FT

XX FT

XX FT

Streptomyces grise

Antigenic protein

Group B Streptococ

Haemophilus influ

Human RNA-associat

Orpiniomyces cellui

Human secreted pro

Neocallimastix pat

N. patriciarum xyl

Human p300 ligand

Human TRIDENT tran

Human prothromone c

Human hepatocyte n

Notch hmx3 full le

Human myotonic dys

Human secreted pro

Hydra head activat

Transcription fact

Cellular transcrip

Amino acid sequenc

Human breast tumou

Human TSC-22 homol

Protein LP40, a va

Sequence encoded b

Human Notch signal

Human Notch signal

Human alpha-6(TV)

Porphyrinomas ging

Human Notch signal

Human Notch signal

Human Notch signal

Truncated HSVB po

HSV-2 strain SB5 C

PN W09516709-A2.
 XX 22-JUN-1995.
 XX 13-DEC-1994; 94MO-US14643.
 XX 23-DEC-1993; 93US-0173512.
 XX 13-DEC-1993; 93US-0166333.
 PA (BIOI) BIOGEN INC.
 PA (INRM) INSERM INST NAT SANTE & RECH MED.
 PI Cate RL, Jossso N;
 DR WPI: 1995-231521/30.
 DR N-PSDB: Q98988.
 XX
 XX New DNA encoding anti-Mullerian hormone receptor, related
 PT polypeptide(s) and antibodies - useful in immunosays screening for
 PT ligands, treatment of cancer cells expressing the receptor etc.

PS Claim 4, Page 54-56; 77pp: English.

XX This sequence represents the protein encoded by the cDNA clone 3-2. This
 CC sequence was found to be the human AMH-receptor. AMH is a glycoprotein
 CC and is part of the transforming growth factor-beta superfamily. AMH is
 CC structurally and functionally related to TGF-beta home morphogenic
 CC protein, and activin. AMH receptors are present in a limited number of
 CC tissues and can therefore be used to design antibody-toxin complexes to
 CC target tumor cells in these tissues. The AMH-receptor sequences can be
 CC used to screen and purify compounds that bind to them, thereby
 CC identifying AMH ligands. The antibody designed from the levels of
 CC sequences can be used in immunosays to detect the levels of
 CC AMH-receptor. The antibody can also be linked to a toxin in order to
 CC kill cells that express the receptor, e.g. cancer cells. The advantage
 CC of using antibody conjugates based on AMH-receptors are that they can be
 CC used more aggressively than those based on antibodies that recognise
 CC receptors more generally distributed.

XX Sequence 573 AA;

Query Match 100.0%; Score 35; DB 16; Length 573;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTP0PG 6
 |||||
 DB 369 ptpgpg 374

RESULT 2
 Y36955
 ID Y36955 standard; Protein; 797 AA.

XX Y36955;

DT 07-OCT-1999 (first entry)

DE Chlamydia trachomatis surface exposed protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perithenaritis;
 KW nongonococcal urethritis; epididymitis; salpingitis;
 KW bartholinitis; pneumonia; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

PN W09928475-A2.

PD 10-JUN-1999.

PF 27-NOV-1998; 98MO-TB01939.

XX 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.

XX (GEST) GENSET.

XX Griffais R;

DR WPI: 1999-371125/31.

PT Genome sequence of Chlamydia trachomatis

PS Disclosure: Page 802; 1755pp; English.

XX Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
 CC Chlamydia trachomatis (see Z01425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perithenaritis, bartholinitis; pneumonia; lymphogranulomatosis;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX Sequence 797 AA;

Query Match 91.4%; Score 32; DB 20; Length 797;
 Best Local Similarity 83.3%; Pred. No. 2,8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTP0PG 6
 ||:|||
 DB 697 ptpgpg 702

RESULT 3
 Y84616
 ID Y84616 standard; Protein; 300 AA.

XX Y84616;

DT 25-JUL-2000 (first entry)

DE A chicken motor neuron restricted pattern 2 protein.

XX Motor neuron restricted pattern 2 protein; MNR2 protein; motor neuron;
 KW Pax6+ motor neuron progenitor; somatic motor neuron;
 KW transcription factor; Is12; Uln 3; homeobox protein 9; HB9;
 KW chronic neurodegenerative disease; spinal muscular atrophy;
 KW myotrophic lateral sclerosis; neuromuscular disease; sacral agenesis.

OS Gallus sp.

PN W0200018884-A1.

PD 06-APR-2000.

PF 29-SEP-1999; 99MO-US22517.

PR 29-SEP-1998; 98US-0162524.

PA (UNCO) UNIV COLUMBIA NEW YORK.

PI Jessell TM, Tanabe Y, William C;

DR WPI: 2000-293133/25.

DR N-PSDB: Z99988.

PT New isolated nucleic acid molecules encoding motor neuron restricted

PT pattern protein and a homeobox protein are used in treating chronic
 PT neurodegenerative diseases and acute nervous system injuries -
 XX
 PS Claim 31; Fig 10; 209pp; English.
 XX
 CC The present sequence represents a motor neuron restricted pattern (MNR2)
 CC protein from a chicken embryo. MNR2 is expressed selectively by Pax6+
 CC motor neuron progenitors and persists in post-mitotic somatic motor
 CC neurons. Expression of the MNR2 protein induces expression of
 CC transcription factors Isl2, Isl3 and homeobox protein 9 (Hb9).
 CC MNR2 and Hb9 are used to induce differentiation of somatic motor
 CC neurons in a mammal which is to treat an abnormality associated with a
 CC lack of one or more normally functioning motor neurons, a chronic
 CC neurodegenerative disease such as spinal muscular atrophy or myotrophic
 CC lateral sclerosis, an acute nervous system injury localized to a
 CC specific central axon and neuromuscular disease. Hb9 is also used to
 CC treat an embryo afflicted with sacral agenesis.
 XX
 SQ Sequence 300 Aa;
 Query Match 88.6%; Score 31; DB 21; Length 300;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PTQPOG 6
 ID 1:|||||
 DB 238 pspgpg 243
 RESULT 4
 W14532
 ID W14532 standard; Protein; 350 AA.
 AC W14532;
 XX
 DT 19-JUN-1997 (first entry)
 XX
 DE Human chimeric fucosyltransferase Fuc-TC21.
 XX
 KW Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TIII;
 KW alpha(1,3)-fucosyltransferase; Fuc-TV; glycosylation;
 KW oligosaccharide; Fuc-TC21.
 XX
 OS Synthetic.
 XX
 FH Key
 FT 1..99 Location/Qualifiers
 FT Region //label= Fuc-TIII
 FT //note= "amino acids 1-99 of Fuc-TIII"
 FT Region 100..150
 FT //label= Fuc-TV
 FT //note= "amino acids 116-166 of Fuc-TV"
 FT Region 151..351
 FT //label= Fuc-TIII
 FT //note= "amino acids 161-361 of Fuc-TIII"
 FT
 XX W09709421-A1.
 PN
 XX 13-MAR-1997.
 PD
 XX 06-SEP-1996; 96WO-US13816.
 PF
 XX 08-SEP-1995; 95DS-0525058.
 PR
 XX (UNMI) UNIV MICHIGAN.
 PA
 XX Legault DJ, Lowe JB;
 PI
 XX WPI: 1997-192897/17.
 DR
 XX New recombinant fucosyltransferase proteins - useful for modifying
 PT cell surface oligosaccharide structures

XX
 PS Claim 1; Refer to Page 298; 329pp; English.
 XX
 CC Chimeric fucosyltransferases Fuc-TC1 to Fuc-TC21 (W14512-32)
 CC comprise portions of the human GDP-Fuc:(beta-D-Gal(1,4/1,3))-D-
 CC GlcNAc(5/6) alpha(1,3/1,4)-Fucosyltransferase (Fuc-TIII, Lewis
 CC enzyme) (W13638) and portions of the GDP-Fuc:(beta-D-Gal(1,4)-D-
 CC GlcNAc alpha(1,3))fucosyltransferase (Fuc-TVI) (W13643) or Fuc-TV
 CC (W13642). They are obid. by cassette mutagenesis of Fuc-TIII and
 CC Fuc-TVI or Fuc-TV nucleic acids (T61675, T61679-80) and expression
 CC in transformed host cells. The chimeric fucosyltransferases can
 CC be used to modify cell surface oligosaccharide structures.
 XX
 SQ Sequence 350 Aa;
 Query Match 88.6%; Score 31; DB 18; Length 350;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PTQPOG 6
 DB 120 plrpg 125
 RESULT 5
 R45939
 ID R45939 standard; Protein; 374 AA.
 AC R45939;
 XX
 DT 26-JUL-1994 (first entry)
 XX
 DE A glycosyltransferase.
 XX
 KW Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
 KW surface; oligosaccharide.
 XX
 OS Homo sapiens.
 XX
 PN W09402616-A.
 XX
 PD 03-FEB-1994.
 XX
 PF 20-JUL-1993; 93WO-US06703.
 XX
 PR 20-JUL-1992; 92US-0914281.
 XX
 PA (UNMI) UNIV MICHIGAN.
 XX
 PI Lowe JB;
 XX
 DR WPI: 1994-048674/06.
 DR N-PSDB; Q56911.
 XX
 PT DNA fragment encoding a glycosyltransferase - can be used for in
 PT vitro reactions to modify cell surface oligosaccharide(s) e.g.
 PT blood gp. determinants, to protect against transplant rejection
 XX
 PS Disclosure; Fig 6; 249pp; English.
 XX
 CC The sequence is that of a human glycosyl transferase. The enzyme
 CC may be non glycosylated. This prevents premature loss of enzyme
 CC activity. It can also be used in in vitro reactions to modify cell
 CC surface oligosaccharide mols. e.g. blood group determinants.
 CC See also R45933-8.
 CC
 XX
 SQ Sequence 374 Aa;
 Query Match 88.6%; Score 31; DB 15; Length 374;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
 DB 137 pcrp9 142

RESULT 6
 W13642

W13642 standard: Protein: 374 AA.

W13642:

19-JUN-1997 (first entry)

Human alpha(1,3)-Fucosyltransferase (Fuc-TV).

Alpha(1,3)-Fucosyltransferase; Lewis enzyme; Fuc-TV;
 glycosylation; oligosaccharide; blood group.

OS Homo sapiens.

Key Location/Qualifiers

FT Domain 16..34

FT Domain /label= Transmembrane_domain

FT Domain 43..374

FT Domain /label= Catalytic_domain

PN W09709421-A1.

PD 13-MAR-1997.

PF 06-SEP-1996: 96WO-US13816.

PR 08-SEP-1995: 95US-0525058.

PA (UNMI) UNIV MICHIGAN.

PI Legault DJ, Lowe JB;

DR WPI: 1997-192897/17.

DR N-PSDB: T61679.

PT Now recombinant fucosyltransferase proteins - useful for modifying
 cell surface oligosaccharide structures

PS Example 5: Page 288-289; 329pp; English.

CC Human GDP-Fuc:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-fucosyltransferase
 CC (Fuc-TV) (W13642) can utilise neutral type II oligosaccharide
 CC precursors but cannot efficiently utilise type I oligosaccharides.
 CC Its amino acid sequence was deduced from an isolated DNA sequence.
 CC obtd. from a human genomic DNA using Lewis enzyme cDNA (see also
 CC T61675) as probe. The Fuc-TV enzyme, when expressed by the cloned
 CC DNA sequence, functions within mammalian cells to generate de novo
 CC expression of specific cell surface glycoconjugate structures that
 CC are recognised by antibodies against SSEA-1, Lewis x or sialyl
 CC Lewis x.

CC Sequence 374 AA;

OY 1 PROPOG 6
 DB 137 pcrp9 142

Query Match 88.6%; Score 31; DB 18; Length 374;
 Best Local Similarity 83.3%; Pred. No. 26+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
 Y20992
 ID Y20992 standard: Protein: 20 AA.

XX Y20992:

XX 22-JUL-1999 (first entry)

XX Human glial fibrillary acidic protein GFAP mutant fragment 1.

XX Human: beta-amyloid precursor protein; beta-Ap; diagnosis; cancer;
 XX frameshift mutation; age-related disease; neurodegenerative disorder;
 XX Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
 XX Huntington's disease; multiple sclerosis; alcoholic liver disease;
 XX diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
 XX ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
 XX neurofilament-F; presenilin 1; presenilin II; cellular tumour antigen;
 XX glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
 XX bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMG-C; NSP-A;
 XX high mobility group protein-C; neuroendocrine specific protein A.

OS Synthetic.

OS Homo sapiens.

PN W09845322-A2.

PD 15-OCT-1998.

PF 02-APR-1998: 98WO-1B00705.

PR 10-APR-1997: 97US-0043163.

PA (UYUT-) RIJKSUNIV UTRECHT.

PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.

PA (UYRO-) UNIV ROTTERDAM ERASMUS.

PI Butbach JPH, Grosfeld FG, Van Leeuwen FW;

DR WPI: 1998-609901/51.

DR N-PSDB: X75764.

PT Diagnosing disease by detecting frameshift mutations in RNA or
 PT corresponding protein mutations - used to diagnose cancer and
 PT neurological diseases, particularly Alzheimer's disease, and also
 PT for treatment and prevention with specific ribozymes or wild-type
 PT RNA

PS Disclosure: Figure 13; 258pp; English.

CC This invention describes a novel method for the diagnosis of a disease
 CC caused by, or associated with, an RNA molecule that has a frameshift
 CC mutation. The method is used to diagnose age-related diseases, especially
 CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
 CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
 CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
 CC and many others listed) or susceptibility to these disorders. The method
 CC allows a definitive diagnosis of Alzheimer's disease in living patients,
 CC at an early stage. It is based on the observation that disease may be
 CC caused by mutations in RNA rather than DNA. The invention describes the
 CC use of neuronal system RNA molecules, specifically proteins including
 CC beta-amyloid precursor protein (beta-Ap), the microtubule associated
 CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
 CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
 CC neurofilament-F, presenilin 1, presenilin II, glial fibrillary acidic
 CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
 CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
 CC protein-C (HMG-C) and neuroendocrine specific protein A.

CC Sequence 20 AA;

OY 1 PROPOG 6
 Query Match 85.7%; Score 30; DB 19; Length 20;
 Best Local Similarity 83.3%; Pred. No. 16;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTP0G 6
DB 20 ptpqg 25

RESULT 10

ID P50761 standard; Protein: 74 AA.

AC P50761;

DT 08-NOV-1991 (first entry)

DE Antigenic protein encoded by clone lambda95.1(8).

KW Malaria; vaccine; antigen; epitope.

OS Plasmodium falciparum K1 isolate (Thailand).

FI Key Location/Qualifiers

FT Peptide 32..49

FT /label= epitope 5.1-1

FT /note= "repeat; see comments"

PN EP153180-A.

PD 28-AUG-1985.

PP 19-FEB-1985; 85EP-0301110.

PR 21-DEC-1984; 84GB-0032337.

PR 21-FEB-1984; 84GB-0004493.

PR 13-NOV-1984; 84GB-0028643.

PA (NATR) NATIONAL RES DEV CORP.

PI Hope IA, Scaife JG, Strambachova-WCBride J;

DR WPI: 1985-211745/35.

DR N-PSDB; N50553.

PT Antigenic material for vaccination against malaria - comprises

PT amino acid residues sequence of natural antigen and

PT intra-erythrocytic forms of plasmodium parasites.

PS Disclosure: Fig 5a; 70pp; English.

CC The sequence was deduced from a DNA insert from clone

CC lambda95.1(8) which was prep. from RNA from the erythrocytic

CC stages of the parasite. The antigen is present as a sporozoite

CC surface epitope. The 5.1-1 epitope comprises a sequence homologous

CC with the tandemly repeated tetramer present in the circumsporozoite

CC protein (MANP). The 5.1-1 epitope appears to be made up of two

CC sequences of seven and eight residues respectively linked by three

CC residues, Ser-Glu-Ser, which do not contribute to antigenicity and

CC may be omitted from synthetic antigens. The DNA can be used to

CC express antigenic peptides useful for the prep. of vaccines for

CC protection against malaria. See also P51201.

CC Sequence 74 AA;

QY 1 PTP0G 6

DB 58 ptpqg 63

Query Match 82.9%; Score 29; DB 6; Length 74;

Best Local Similarity 83.3%; Pred. No. 91;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

RESULT 11

Y21354

ID Y21354 standard; Protein: 125 AA.

AC Y21354;

DT 22-JUL-1999 (first entry)

DE Human HUPF-I mutant protein fragment 6.

KW Human: beta-amyloid precursor protein; beta-APP; diagnosis; cancer;

KW frameshift mutation; age-related disease; neurodegenerative disorder;

KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;

KW Huntington's disease; multiple sclerosis; alcoholic liver disease;

KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;

KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;

KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;

KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;

KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;

KW high mobility group protein-C; neuroendocrine specific protein A.

OS Synthetic.

OS Homo sapiens.

PN W09845322-A2.

PD 15-OCT-1998.

PP 02-APR-1998; 98MO-1B00705.

PR 10-APR-1997; 97US-0043163.

PA (UYUT-) RIKSUNIV UTRECHT.

PA (ROYA-) ROYAL NETHERLANDS ACADEMIES OF SCI.

PA (UYRO-) UNIV ROTTERDAM ERASMUS.

PI Burbach JPH, Grosveld FG, Van Leeuwen FW;

DR WPI: 1998-609901/51.

DR N-PSDB; X75768.

PT Diagnosing disease by detecting frameshift mutations in RNA or

PT corresponding protein mutations - used to diagnose cancer and

PT neurological diseases, particularly Alzheimer's disease, and also

PT for treatment and prevention with specific ribozymes or wild-type

PS Disclosure: Figure 17; 258pp; English.

CC This invention describes a novel method for the diagnosis of a disease

CC caused by, or associated with, an RNA molecule that has a frameshift

CC mutation. The method is used to diagnose age-related diseases, especially

CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's

CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,

CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II

CC and many others listed) or susceptibility to these disorders. The method

CC allows a definitive diagnosis of Alzheimer's disease in living patients,

CC at an early stage. It is based on the observation that disease may be

CC caused by mutations in RNA rather than DNA. The invention describes the

CC use of neuronal system RNA molecules, specifically proteins including the

CC beta-amyloid precursor protein (beta-APP), the microtubule associated

CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule

CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,

CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic

CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma

CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group

CC protein-C (HMGP-C) and neuroendocrine specific protein A.

CC Sequence 125 AA;

QY 1 PTP0G 6

DB 58 ptpqg 63

Query Match 82.9%; Score 29; DB 19; Length 125;

Best Local Similarity 83.3%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 PROPOG 6
11111
Db 14 pepqpg 19

RESULT 12
ID Y73845
Y73845 standard; Protein: 127 AA.

AC Y73845:
XX
XX 14-MAR-2000 (first entry)
DT

DE Human prostate tumor EST fragment derived protein #32.
XX
XX Pancreas; tumor; EST, expressed sequence tag; human; cytostatic;
XX treatment.

XX OS Homo sapiens.
XX
XX DE19820190-A1.
PN

XX PD 04-NOV-1999.

XX PF 28-APR-1998; 98DE-1020190.

XX PR 28-APR-1998; 98DE-1020190.

XX (META-) METAGEN CBS GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI: 1999-621386/54.

XX DR N-PSDB; 252868.

XX PT New human nuclear acid sequences from pancreatic tumors, and related
XX proteins

XX PS Claim 23; Page 371; 502pp; German.

XX CC This invention describes novel polypeptides and their encoding nucleic
XX CC activity derived from human pancreatic tumor tissue which have cytostatic
XX CC activity. The sequences are also useful in producing pharmaceutical
XX CC compositions for treatment of pancreatic tumors. Y73814-Y74257 represent
XX CC protein fragments encoded by the human pancreatic tumor cDNA library
XX CC derived expressed sequence tag (EST) sequences represented in
XX CC 252868-253014.

XX SQ Sequence 127 AA;

Query Match 82.9%; Score 29; DB 20; Length 127;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PROPOG 6
11111
Db 18 pepqpg 23

RESULT 13
ID Y44649
Y44649 standard; Protein: 135 AA.

AC Y44649:
XX
XX 18-APR-2000 (first entry)
DT

XX DE Streptomyces griseus SsgA protein.

XX KW SsgA; liquid culture; filamentous bacteria; secondary metabolite;
XX KW mycelium; antibiotic; antitumor agent; immunosuppressive agent;
XX KW hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;

KW herbicide; antiparasitic agent; ruminant growth promoter;
KW bioinsecticide; receptor agonist; antagonist; biomass.
XX
XX Streptomyces griseus.
XX OS

XX PN WO200000613-A1.

XX PD 06-JAN-2000.

XX PF 25-JUN-1999; 99MO-NL00395.

XX PR 26-JUN-1998; 98EP-0202148.

XX PA (UYLE-) RIJKSUNIV LEIDEN.
XX (NEME-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZ.

XX PI Van Wezel GP, Kraal B, Luiten RGM;

XX DR WPI: 2000-147269/13.

XX DR N-PSDB; 249727, 249728.

XX PT Reducing branching and enhancing fragmentation in filamentous
XX microorganisms used to improve their liquid culturing properties

XX PS Disclosure; Fig 5; 60pp; English.

XX CC The present sequence is S. griseus SsgA protein. SsgA reduces branching
XX CC and fragment septation and enhances fragmentation of mycelium in liquid
XX CC culture resulting in lower viscosity of culture broths. Filamentous
XX CC bacteria can be transformed with ssgA gene-containing plasmid to enhance
XX CC the production of secondary metabolites such as, antibiotics, antitumor
XX CC agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme
XX CC inhibitors, antimigraine agents, herbicides, antiparasitic agents,
XX CC ruminant growth promoters, bioinsecticides, receptor agonists and
XX CC antagonists and biomass.

XX SQ Sequence 135 AA;

Query Match 82.9%; Score 29; DB 21; Length 135;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
11111
Db 65 ptepg 70

RESULT 14
ID P51201
P51201 standard; Protein: 162 AA.

AC P51201:
XX
XX 08-NOV-1991 (first entry)
DT

XX DE Antigenic protein encoded by clone lambdaDAg5.1(9).

XX KW Malaria; vaccine; antigen; epitope.

XX OS Plasmodium falciparum KI isolate (Thailand).

XX FH Key Location/Qualifiers

XX FT Peptide 1..22 /label= signal sequence

XX FT Protein 23..162 /label= mature protein

XX FT Peptide 120..137 /label= epitope 5.1-1

XX FT /note= "repeat; see comments"

XX FT Region 80..101 /label= stop transfer sequence

XX FT /note= "hydrophobic"

XX EPI53188-A.
 XX 28-AUG-1985.
 PD
 XX 19-FEB-1985; 85EP-0301110.
 PF
 XX 21-DEC-1984; 84GB-0032337.
 PR 21-FEB-1984; 84GB-0004493.
 PR 13-NOV-1984; 84GB-0028643.
 XX
 PA (NATR) NATIONAL RES DEV CORP.
 PI Hope IA, Scaife JG, Strambachova-McBride J;
 XX WPI: 1985-211745/35.
 DR N-PSDB: N50851.
 XX
 PT Antigenic material for vaccination against malaria - comprises
 PT amino acid residues sequence of natural antigen and
 PT intra-erythrocytic forms of plasmodium parasites.
 XX
 PS Claim 13: Fig 5a; 70pp; English.
 XX
 CC The sequence was deduced from a DNA insert from clone lambda95.1(9)
 CC which was prep'd. from RNA from the erythrocytic stages of the
 CC parasite. The antigen is present as a sporozoite surface epitope.
 CC The 5.1-1 epitope comprises a sequence homologous with the tandemly
 CC repeated tetramer present in the circumsporozoite protein (NMP).
 CC The 5.1-1 epitope appears to be made up of two sequences of seven
 CC and eight residues respectively linked by three residues. Ser-Glu-
 CC Ser, which do not contribute to antigenicity and may be omitted
 CC from synthetic antigens. The DNA can be used to express antigenic
 CC peptides useful for the prep'n. of vaccines for protection against
 CC malaria. See also p50761.
 XX
 SQ Sequence 162 AA;

Query Match 82.9%; Score 29; DB 6; Length 162;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PROPG 6
 Db 146 pcpqg 151

RESULT 15
 Y91338
 ID Y91338 standard; Protein; 262 AA.
 XX
 AC Y91338;
 XX
 XX
 DT 30-MAY-2000 (first entry)
 XX
 DE Group B Streptococcus protein sequence SEQ ID NO:71.
 XX
 KW Group B Streptococcus; Streptococcus agalactiae; protein antigen;
 KM vaccine; screening; immunogen; detection; diagnosis; infection;
 XX antibody; aflibody; antibacterial.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200006736-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB02444.
 XX
 PR 27-JUL-1999; 98GB-0016335.
 PR 19-MAR-1999; 99US-0125163.
 XX

PA (MICR-) MICROBIAL TECHNICS LTD.
 XX
 PI Le Page RWF, Wells JM, Hanniffy SB;
 XX
 DR WPI: 2000-195299/17.
 XX
 PT New Group B Streptococcus protein, useful as vaccine, for diagnosis of
 PT Streptococcal infections and for screening of antibodies or aflibodies
 XX
 PS Claim 1; Fig 1; 123pp; English.
 XX
 CC A05803 to A05872 encode proteins, polypeptides and peptides (given in
 CC Y91275 to Y91343) isolated from Group B Streptococcus (GBS), also known
 CC as Streptococcus agalactiae. The GBS polynucleotides and polypeptides
 CC have antibacterial activity. Immunogenic compositions comprising GBS
 CC polynucleotides or polypeptides can be used as vaccines and for the
 CC treatment or prophylaxis of GBS infection. The polynucleotides and
 CC polypeptides can also be used in the detection of GBS and for screening
 CC DNA encoding bacterial cell envelope associated or secreted antigens in
 CC gram positive bacteria. A05873 to A05941 represent primers used in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 262 AA;

Query Match 82.9%; Score 29; DB 21; Length 262;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPG 5
 Db 26 pcpqg 30

Search completed: November 16, 2000, 15:36:47
 Job time: 37 sec

•

•

•

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 16, 2000, 15:36:09 ; Search time 12.25 Seconds
(without alignments)
8,210 Million cell updates/sec

Title: US-09-308-237B-1

Perfect score: 35

Sequence: 1 PRTQPG 6

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cyn2.6/pdata/2/1aa/5A.COMB.pep:*
2: /cyn2.6/pdata/2/1aa/5B.COMB.pep:*
3: /cyn2.6/pdata/2/1aa/6.COMB.pep:*
4: /cyn2.6/pdata/2/1aa/PCCTS.COMB.pep:*
5: /cyn2.6/pdata/2/1aa/backfilltest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	31	88.6	374	1 US-07-914-281-11	Sequence 11, Appl
2	31	88.6	374	1 US-08-393-246-11	Sequence 11, Appl
3	31	88.6	374	1 US-08-525-058A-11	Sequence 11, Appl
4	31	88.6	374	2 US-08-696-731-11	Sequence 11, Appl
5	29	82.9	162	2 US-08-319-704-6	Sequence 6, Appl
6	29	82.9	327	2 US-08-651-818A-3	Sequence 3, Appl
7	29	82.9	459	3 US-09-118-319-6	Sequence 6, Appl
8	29	82.9	485	2 US-08-749-391-2	Sequence 2, Appl
9	29	82.9	516	2 US-08-676-166A-7	Sequence 7, Appl
10	29	82.9	755	3 US-09-071-101-2	Sequence 2, Appl
11	29	82.9	755	3 US-09-369-618-2	Sequence 2, Appl
12	29	82.9	763	2 US-08-742-753-4	Sequence 4, Appl
13	29	82.9	1064	1 US-08-537-210A-3	Sequence 3, Appl
14	29	82.9	1068	1 US-08-537-210A-2	Sequence 2, Appl
15	29	82.9	1078	1 US-08-264-534-32	Sequence 32, Appl
16	29	82.9	1078	1 US-08-083-590A-11	Sequence 11, Appl
17	29	82.9	1078	1 US-08-465-500-32	Sequence 32, Appl
18	29	82.9	1078	3 US-08-346-128-32	Sequence 32, Appl
19	29	82.9	1078	3 US-08-532-384-11	Sequence 11, Appl
20	29	82.9	1078	3 US-08-893-828-32	Sequence 32, Appl
21	29	82.9	2414	4 US-08-227-536-2	Sequence 2, Appl
22	29	82.9	2414	4 PCT-US85-04682-2	Sequence 2, Appl
23	29	82.9	2523	1 US-08-185-432-18	Sequence 18, Appl
24	29	82.9	2556	1 US-08-185-432-11	Sequence 11, Appl
25	29	82.9	2556	1 US-08-083-590A-20	Sequence 20, Appl
26	29	82.9	2556	1 US-08-532-384-20	Sequence 20, Appl
27	28	80.0	143	2 US-08-889-337-3	Sequence 3, Appl
28	28	80.0	143	3 US-08-851-190-6	Sequence 6, Appl

29	28	80.0	144	2 US-08-889-337-1	Sequence 1, Appl
30	28	80.0	144	3 US-08-851-190-7	Sequence 7, Appl
31	28	80.0	547	1 US-08-494-168-7	Sequence 7, Appl
32	28	80.0	582	3 US-08-906-865-3	Sequence 3, Appl
33	28	80.0	903	1 US-08-220-151-8	Sequence 8, Appl
34	28	80.0	903	1 US-08-413-118-8	Sequence 8, Appl
35	28	80.0	903	3 US-08-804-439A-22	Sequence 22, Appl
36	28	80.0	903	3 US-08-473-446-8	Sequence 8, Appl
37	28	80.0	903	3 US-08-720-229-22	Sequence 22, Appl
38	28	80.0	904	4 PCT-US96-05316-1	Sequence 1, Appl
39	28	80.0	904	4 PCT-US96-05316-2	Sequence 2, Appl
40	28	80.0	904	5 5244792-3	Patent No. 5244792
41	28	80.0	904	5 5244792-4	Patent No. 5244792
42	28	80.0	1243	2 US-08-557-139-2	Sequence 2, Appl
43	28	80.0	1694	1 US-08-494-168-2	Sequence 2, Appl
44	26	74.3	105	1 US-08-417-460-6	Sequence 6, Appl
45	26	74.3	301	4 PCT-US95-13975-72	Sequence 72, Appl

ALIGNMENTS

RESULT 1
US-07-914-281-11
; Sequence 11, Application US/07914281
; Patent No. 5324663
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOPOLYMERIS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/914,281
; FILING DATE: 19920720
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELETYPE: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-07-914-281-11

Query Match 88.6%; Score 31; DB 1; length 374;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0;
Gaps 0;
OY 1 PRTQPG 6
|||||

Db 137 PTRPOG 142

RESULT 2

US-08-393-246-11
Sequence 11, Application US/08393246
Patent No. 5595900

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
NUMBER OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURE

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,

ADDRESS: P. C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/393,246

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/220,433

FILING DATE: 30-MAR-1994

APPLICATION NUMBER: US 07/914,281

FILING DATE: 20-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Lavelle, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 374 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-393-246-11

Query Match 88.6%; Score 31; DB 1; Length 374;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTRPOG 6

Db 137 PTRPOG 142

RESULT 3
US-08-525-058A-11
Sequence 11, Application US/08525058A
Patent No. 5770420

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
NUMBER OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,

ADDRESS: P. C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/525,058A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/393,246

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT, P. C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/525,058A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lavelle, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 374 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-525-058A-11

Query Match 88.6%; Score 31; DB 1; Length 374;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTRPOG 6

Db 137 PTRPOG 142

RESULT 4

US-08-696-731-11

Sequence 11, Application US/08696731

Patent No. 5955347

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
NUMBER OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURE

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,

ADDRESS: P. C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/696,731

FILING DATE: 14-AUG-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/393,246

APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavallee, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248955 OPAT UR
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-696-731-11

Query Match 88.6% Score 31; DB 2; Length 374;
Best Local Similarity 83.3% Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPQPG 6
111111
Db 137 PTPQPG 142

RESULT 5
US-08-319-704-6
Sequence 6, Application US/08319704
Patent No. 5814617
GENERAL INFORMATION:
APPLICANT: Hoffman, Stephen L.
APPLICANT: Charcenvit, Yudin
APPLICANT: Hedstrom, Richard C.
APPLICANT: Doolan, Denise L.
TITLE OF INVENTION: Protective 17 kDa Malaria Hepatic and
TITLE OF INVENTION: Erythrocytic Stage Immunogen and Gene
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Naval Medical R & D Command
STREET: Bldg 1, T-12, 8901 Wisconsin Avenue
CITY: Bethesda
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,704
FILING DATE: 07-OCT-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: A. David Spevack
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-1022
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-319-704-6

Query Match 82.9% Score 29; DB 2; Length 162;
Best Local Similarity 83.3% Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTPQPG 6
111111
Db 146 PTPQPG 151

RESULT 6
US-08-651-818A-3
Sequence 3, Application US/08651818A
Patent No. 5948889
GENERAL INFORMATION:
APPLICANT: de Boer, Piet A.J.
APPLICANT: Hale, Cynthia A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING
TITLE OF INVENTION: ANTIMICROBIALS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,818A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02249
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-651-818A-3

Query Match 82.9% Score 29; DB 2; Length 327;
Best Local Similarity 100.0% Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPQPG 5
111111
Db 123 PTPQPG 127

RESULT 7
US-09-118-319-6
Sequence 6, Application US/09118319
Patent No. 6114158
GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Chen, Huizhong
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Ornithomycetes Cellulase Celf Protein and Coding Sequences
FILE REFERENCE: 33-98sequence listing
CURRENT APPLICATION NUMBER: US/09/118,319

;; CURRENT FILING DATE: 1998-07-17
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 6
;; LENGTH: 459
;; TYPE: PRT
;; ORGANISM: Orplomyces sp. PC-2
US-09-118-319-6

Query Match 82.9%; Score 29; DB 3; Length 459;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PTOPOG 5
| | | | |
Db 119 PTOPOG 123

RESULT 8
US-08-749-391-2
; Sequence 2, Application US/08749391
; Patent No. 5948667
; GENERAL INFORMATION:
; APPLICANT: Cheng, Kuo-Jean
; APPLICANT: Selinger, Leonard B.
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Hu, Youji
; APPLICANT: Forsberg, Cecil W.
; TITLE OF INVENTION: A xylanase Obtained From an
; TITLE OF INVENTION: Anaerobic Fungus
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80803
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,391
; FILING DATE: 13-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna M. Ferber
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 93-96
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-749-391-2

Query Match 82.9%; Score 29; DB 2; Length 485;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTOPOG 6
| | | | |
Db 331 PTOPOG 336

RESULT 9
US-08-676-166A-7
; Sequence 7, Application US/08676166A
; Patent No. 5953270
; GENERAL INFORMATION:
; APPLICANT: Radford, Alan
; APPLICANT: Parish, John H.
; TITLE OF INVENTION: EXPLOITATION OF THE CELLULOSE COMPLEX OF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,166A
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: P. chrysosporium
US-08-676-166A-7

Query Match 82.9%; Score 29; DB 2; Length 516;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTOPOG 6
| | | | |
Db 474 PTOPOG 479

RESULT 10
US-09-071-101-2
; Sequence 2, Application US/09071101
; Patent No. 6013503
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Taspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.101
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E.
REGISTRATION NUMBER: 31,648
REFERENCE/DOCKET NUMBER: 57-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 755 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-09-071-101-2

Query Match 82.9%; Score 29; DB 3; Length 755;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTOPO 5
Db 448 PTOPO 452

RESULT 11
US-09-369-618-2
Sequence 2, Application US/09369618
Patent No. 6100041
GENERAL INFORMATION:
APPLICANT: Lok, Si
TITLE OF INVENTION: HUMAN PROTEINOME CONVERTASE 4
FILE REFERENCE: 97-05D2
CURRENT APPLICATION NUMBER: US/09/369,618
CURRENT FILING DATE: 1999-08-06
EARLIER APPLICATION NUMBER: US 09/071,101
EARLIER FILING DATE: 1998-05-01
EARLIER APPLICATION NUMBER: US 60/044,015
EARLIER FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 755
TYPE: PRT
ORGANISM: Homo sapiens
US-09-369-618-2

Query Match 82.9%; Score 29; DB 3; Length 755;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTOPO 5
Db 448 PTOPO 452

RESULT 12
US-08-742-753-4
Sequence 4, Application US/08742753
Patent No. 5861278
GENERAL INFORMATION:
APPLICANT: WONG, Gordon G.
TITLE OF INVENTION: HNF3-delta Compositions
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,753
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5277
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8260
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 70..2358
US-08-742-753-4

Query Match 82.9%; Score 29; DB 2; Length 763;
Best Local Similarity 83.3%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTOPO 6
Db 115 PTOPO 120

RESULT 13
US-08-537-210A-3
Sequence 3, Application US/08537210A
Patent No. 5780300
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,210A
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1064 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: unknown
FEATURE:
NAME/KEY: Xen N
LOCATION: 1150...2213
OTHER INFORMATION: Highly conserved ankyrin repeat
US-08-537-210A-3

```

```

Query Match      82.9%; Score 29; DB 1; Length 1064;
Best Local Similarity 83.3%; Pred. No. 8.1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PTOPOG 6
DB 706 PTPPOG 711

```

```

RESULT 14
US-08-537-210A-2
Sequence 2, Application US/08537210A
Patent No. 5780300
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MAINTENANCE OF NON-TERMINALLY
NUMBER OF INVENTIONS: 4
DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,210A
FILING DATE: 29-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090

```

```

TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1068 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: unknown
FEATURE:
NAME/KEY: Human N1 (TAN-1)
LOCATION: 1152...2219
OTHER INFORMATION: Highly conserved ankyrin repeat
US-08-537-210A-2

```

```

Query Match      82.9%; Score 29; DB 1; Length 1068;
Best Local Similarity 83.3%; Pred. No. 8.2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PTOPOG 6
DB 710 PTPPOG 715

```

```

RESULT 15
US-08-264-534-32
Sequence 32, Application US/08264534
Patent No. 5648464
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
NUMBER OF INVENTIONS: In Topotypic Proteins, And Methods Based Thereon
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/264,534
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/695,189
FILING DATE: 03-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-264-534-32

```

```

Query Match      82.9%; Score 29; DB 1; Length 1078;
Best Local Similarity 83.3%; Pred. No. 8.2e+02;

```

Thu Nov 16 15:44:23 2000

us-09-308-237b-1.ra1

Page 7

Matches	5;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	1	PROPQ6	6						
			11111						
Db	407	PRPQ6	412						

Search completed: November 16, 2000, 15:36:28
Job time: 19 sec

2

3

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 16, 2000, 15:36:10 ; Search time 38.32 seconds
(without alignments)
21.314 Million cell updates/sec

Title: US-09-308-237b-1

Sequence: 1 PROXG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 872601 seqs, 136127174 residues

Total number of hits satisfying chosen parameters: 872601

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents AA:*
1: /cgn2_6/ptodata/2/paa/PCRTUS.COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06.COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07.COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08.COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081.COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082.COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083.COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084.COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US084B.COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US085.COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US086.COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US087.COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US088.COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US089.COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US090.COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US091.COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US092.COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US093.COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US094.COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US095.COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US095.COMB.pep:*
22: /cgn2_6/ptodata/2/paa/PCRTUS.COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US06.COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US07.COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US08.COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US09.COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US09.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	35	100.0	6	US-09-308-237b-1
2	35	100.0	573	PCT-US94-14643-13
3	32	91.4	601	US-60-191-637-21545
4	32	91.4	601	US-60-191-681-16991
5	31	88.6	300	PCT-US99-22517-1
6	31	88.6	300	US-09-162-524-1

7	31	88.6	302	US-09-443-704-12	Sequence 12, Appl
8	31	88.6	374	US-08-823-489-11	Sequence 11, Appl
9	31	88.6	15	US-09-042-531-11	Sequence 11, Appl
10	30	85.7	48	US-60-162-245-2777	Sequence 2777, Ap
11	30	85.7	58	US-60-160-203-5935	Sequence 5935, Ap
12	30	85.7	71	US-60-177-571-2970	Sequence 2970, Ap
13	30	85.7	71	US-60-177-571-2820	Sequence 2820, Ap
14	30	85.7	76	US-60-160-203-4845	Sequence 4845, Ap
15	30	85.7	76	US-60-169-840-6635	Sequence 6635, Ap
16	30	85.7	76	US-60-169-841-2776	Sequence 2776, Ap
17	30	85.7	109	US-60-194-508-3365	Sequence 3365, Ap
18	30	85.7	111	US-60-197-873-22724	Sequence 22724, A
19	30	85.7	119	US-60-181-430-304	Sequence 304, App
20	30	85.7	137	US-60-189-458-157	Sequence 157, App
21	30	85.7	161	US-60-177-571-2750	Sequence 2750, App
22	30	85.7	163	US-60-188-631-141	Sequence 141, App
23	30	85.7	163	US-60-190-000-282	Sequence 282, App
24	30	85.7	174	US-60-189-458-158	Sequence 158, App
25	30	85.7	188	US-09-252-691-5929	Sequence 5929, App
26	30	85.7	188	US-09-417-507-34368	Sequence 34368, A
27	30	85.7	260	US-09-417-507-34912	Sequence 34912, A
28	30	85.7	264	US-09-248-796-19607	Sequence 19607, A
29	30	85.7	264	US-60-096-409-19607	Sequence 19607, A
30	30	85.7	325	US-60-212-656-344	Sequence 344, App
31	30	85.7	325	US-60-230-435-1045	Sequence 1045, App
32	30	85.7	333	US-09-417-507-40369	Sequence 40369, A
33	30	85.7	433	US-60-167-217-15227	Sequence 15227, A
34	30	85.7	433	US-60-173-464-12437	Sequence 12437, A
35	30	85.7	433	US-60-191-637-15191	Sequence 15191, A
36	30	85.7	433	US-60-191-637-15191	Sequence 15191, A
37	30	85.7	795	US-60-191-637-31777	Sequence 31777, A
38	30	85.7	858	US-60-167-217-8669	Sequence 8669, App
39	30	85.7	858	US-60-173-464-6958	Sequence 6958, App
40	30	85.7	858	US-60-191-637-8610	Sequence 8610, App
41	30	85.7	858	US-60-191-681-6693	Sequence 6693, App
42	30	85.7	2632	US-60-230-435-1528	Sequence 1528, App
43	29	82.9	40	US-09-215-026-7	Sequence 7, Appl
44	29	82.9	53	US-60-170-373-2471	Sequence 2471, App
45	29	82.9	65	US-60-171-494-1105	Sequence 1105, App

ALIGNMENTS

RESULT 1
US-09-308-237b-1
Sequence 1, Application US/09308237B
GENERAL INFORMATION:
APPLICANT: OHKI, HIDEKORI
APPLICANT: TOMISHIMA, MASAKI
APPLICANT: YAMADA, AKIRA
APPLICANT: TAKASUGI, HISASHI
TITLE OF INVENTION: CYCLOHEXAPEPTIDES HAVING ANTIMICROBIAL ACTIVITY
FILE REFERENCE: 0018-1052-DPCT
CURRENT APPLICATION NUMBER: US/09/308, 237B
PRIOR APPLICATION NUMBER: PCT/JP97/04193
PRIOR FILING DATE: 1997-11-18
PRIOR APPLICATION NUMBER: AU P03814
PRIOR FILING DATE: 1996-11-25
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 6
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence:peptide
NAME/KEY: MOD_RES
LOCATION: (1)
OTHER INFORMATION: 4-hydroxy
NAME/KEY: MOD_RES
LOCATION: (2)

```

OTHER INFORMATION: hydroxy, substituted phenyl
NAME/KEY: MOD_RES
LOCATION: (3)
OTHER INFORMATION: hydroxy substituted
NAME/KEY: MOD_RES
LOCATION: (4)
OTHER INFORMATION: 3-hydroxy, 4-methyl proline
NAME/KEY: MOD_RES
LOCATION: (5)
OTHER INFORMATION: hydroxy substituted
NAME/KEY: MOD_RES
LOCATION: (6)
OTHER INFORMATION: hydroxy ethyl substituted
OTHER INFORMATION: glycine at position 6 is linked to proline at
OTHER INFORMATION: position 1 by a peptide linkage
US-09-308-237b-1

```

```

Query Match          100.0%; Score 35; DB 26; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 PTPPOG 6
    |||||
DB 1 PTPPOG 6

```

RESULT 2

```

PCT-US94-14643-13
Sequence 13, Application PC/TUS9414643
GENERAL INFORMATION:

```

```

APPLICANT: BIOGEN, INC.
APPLICANT: CARL, Richard L.
APPLICANT: INSERM, (U.293)
APPLICANT: JOSO, Nathalie
TITLE OF INVENTION: ANTI-MULLERIAN HORMONE RECEPTOR
TITLE OF INVENTION: POLYPEPTIDES AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESS: C/O FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14643
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/166,333
FILING DATE: 13-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,512
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B174CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9090
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

```

```

FEATURE:
NAME/KEY: Domain
LOCATION: 18..144
OTHER INFORMATION: /note= "extracellular domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 145..170
OTHER INFORMATION: /note= "transmembrane region"
FEATURE:
NAME/KEY: Domain
LOCATION: 171..573
OTHER INFORMATION: /note= "cytoplasmic domain"
PCT-US94-14643-13

```

```

Query Match          100.0%; Score 35; DB 1; Length 573;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 PTPPOG 6
    |||||
DB 369 PTPPOG 374

```

RESULT 3

```

US-60-191-637-21545
Sequence 21545, Application US/60191637
GENERAL INFORMATION:

```

```

APPLICANT: Ventier, J. Craig
TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
FILE REFERENCE: CL000392
CURRENT APPLICATION NUMBER: US/60/191,637
CURRENT FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 42660
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21545
LENGTH: 601
TYPE: PRT
ORGANISM: DROSOPHILA
US-60-191-637-21545

```

```

Query Match          91.4%; Score 32; DB 21; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 PTPPOG 6
    |||||
DB 338 PTPPOG 343

```

RESULT 4

```

US-60-191-681-16991
Sequence 16991, Application US/60191681
GENERAL INFORMATION:

```

```

APPLICANT: Li, Peter, W.D.
TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
FILE REFERENCE: C1000390
CURRENT APPLICATION NUMBER: US/60/191,681
CURRENT FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 30973
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16991
LENGTH: 601
TYPE: PRT
ORGANISM: DROSOPHILA
US-60-191-681-16991

```

Query Match 91.4%; Score 32; DB 21; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
1:1111
DB 338 PROPOG 343

RESULT 5
PCT-US99-22517-1
Sequence 1, Application PC/TUS9922517

GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University, in the City of
TITLE OF INVENTION: Gene Encoding MNR2 and Uses Thereof

FILE REFERENCE: 57477 A-PCT/jpw/v1

CURRENT FILING DATE: 1999-09-29

EARLIER FILING DATE: 1998-09-29

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentln Ver. 2.0 - beta

SEQ ID NO 1

LENGTH: 300

TYPE: PRT

ORGANISM: chick embryo

PCT-US99-22517-1

Query Match 88.6%; Score 31; DB 1; Length 300;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
1:1111
DB 238 PROPOG 243

RESULT 6
US-09-162-524-1
Sequence 1, Application US/09162524

GENERAL INFORMATION:

APPLICANT: Jessell, Thomas M.

APPLICANT: Tanabe, Yasuo

APPLICANT: William, Christopher

TITLE OF INVENTION: Gene Encoding MNR2 and Uses Thereof

FILE REFERENCE: 57477/jpw/v1

CURRENT FILING DATE: 1998-09-29

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 1

LENGTH: 300

TYPE: PRT

ORGANISM: chick embryo

US-09-162-524-1

Query Match 88.6%; Score 31; DB 16; Length 300;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
1:1111
DB 238 PROPOG 243

RESULT 7
US-09-443-704-12
Sequence 12, Application US/09443704

GENERAL INFORMATION:

APPLICANT: Cahoon, Rebecca E.

APPLICANT: Liu, Zhan-Bin

APPLICANT: Odell, Joan
APPLICANT: Rafalski, Antoni
APPLICANT: Shi, June
APPLICANT: Weng, Zude
TITLE OF INVENTION: Plant MYB-Related Transcription Factors
FILE REFERENCE: B81280 US NA
CURRENT FILING DATE: 1999-11-19
EARLIER FILING DATE: November 20, 1998
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Microsoft Office 97
SEQ ID NO 12
LENGTH: 302
TYPE: PRT
ORGANISM: Glycine max
US-09-443-704-12

Query Match 88.6%; Score 31; DB 18; Length 302;
Best Local Similarity 83.3%; Pred. No. 8.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
1:1111
DB 220 PROPOG 225

RESULT 8
US-08-823-489-11
Sequence 11, Application US/08823489

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURE

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSER: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/823,489

FILING DATE: 25-MAR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/914,281

FILING DATE: 20-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 374 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-823-489-11

Query Match 88.6%; Score 31; DB 13; Length 374;
 Best Local Similarity 83.3%; Pred. No. 1e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 |||||
 Db 137 PTPPOG 142

RESULT 9
 US-09-042-531-11
 ; Sequence 11, Application US/09042531
 ; GENERAL INFORMATION:
 ; APPLICANT: LOWE, JOHN B.
 ; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
 ; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
 ; TITLE OF INVENTION: GLYCOPOLIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
 ; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
 ; ADDRESSEE: P. C.
 ; STREET: 1755 Jefferson Davis Highway, Fourth Floor
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/042,531
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/393,246
 ; FILING DATE:
 ; APPLICATION NUMBER: US 08/220,433
 ; FILING DATE: 30-MAR-1994
 ; APPLICATION NUMBER: US 07/914,281
 ; FILING DATE: 20-JUL-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lavallee, Jean-Paul M. P.
 ; REGISTRATION NUMBER: 31,451
 ; REFERENCE/DOCKET NUMBER: 2363-060-55
 ; TELEPHONE: (703)521-4500
 ; TELEFAX: (703)486-2347
 ; TELEX: 248855 OPAT UR
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 374 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-09-042-531-11

Query Match 88.6%; Score 31; DB 15; Length 374;
 Best Local Similarity 83.3%; Pred. No. 1e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 |||||
 Db 137 PTPPOG 142

RESULT 10
 US-60-162-245-2777

; Sequence 2777, Application US/60162245
 ; GENERAL INFORMATION:
 ; APPLICANT: Bonazzi, Vivien
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: CL0000124
 ; CURRENT APPLICATION NUMBER: US/60/162,245
 ; NUMBER OF SEQ ID NOS: 1999-10-29
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 2777
 ; LENGTH: 48
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-60-162-245-2777

Query Match 85.7%; Score 30; DB 21; Length 48;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 |||||
 Db 41 PTPPOG 46

RESULT 11
 US-60-160-203-5935
 ; Sequence 5935, Application US/60160203
 ; GENERAL INFORMATION:
 ; APPLICANT: BONAZZI, VIVIEN
 ; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
 ; FILE REFERENCE: CL0000116
 ; CURRENT APPLICATION NUMBER: US/60/160,203
 ; CURRENT FILING DATE: 1999-10-19
 ; NUMBER OF SEQ ID NOS: 6374
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 5935
 ; LENGTH: 58
 ; TYPE: PRT
 ; ORGANISM: HUMAN
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)-(58)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 ; US-60-160-203-5935

Query Match 85.7%; Score 30; DB 21; Length 58;
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 |||||
 Db 33 PTPPOG 38

RESULT 12
 US-60-177-571-2970
 ; Sequence 2970, Application US/60177571
 ; GENERAL INFORMATION:
 ; APPLICANT: Bonazzi, Vivien
 ; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
 ; FILE REFERENCE: CL0000201
 ; CURRENT APPLICATION NUMBER: US/60/177,571
 ; CURRENT FILING DATE: 2000-01-27
 ; NUMBER OF SEQ ID NOS: 5082
 ; SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 2970
LENGTH: 58
TYPE: PRT
ORGANISM: HUMAN
US-60-177-571-2970

Query Match 85.7%; Score 30; DB 21; Length 58;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPDPOG 6
11111
Db 16 PTPDPOG 21

RESULT 13
US-60-177-571-2820
Sequence 2820, Application US/60177571
GENERAL INFORMATION:
APPLICANT: Bonazzi, Vivien
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CLC00201
CURRENT APPLICATION NUMBER: US/60/177,571
CURRENT FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 5082
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2820
LENGTH: 71
TYPE: PRT
ORGANISM: HUMAN
US-60-177-571-2820

Query Match 85.7%; Score 30; DB 21; Length 71;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPDPOG 6
11111
Db 34 PTPDPOG 39

RESULT 14
US-60-160-203-4845
Sequence 4845, Application US/60160203
GENERAL INFORMATION:
APPLICANT: BONAZZI, VIVIEN
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: CLO00116
CURRENT APPLICATION NUMBER: US/60/160,203
CURRENT FILING DATE: 1999-10-19
NUMBER OF SEQ ID NOS: 6374
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4845
LENGTH: 76
TYPE: PRT
ORGANISM: HUMAN
US-60-160-203-4845

Query Match 85.7%; Score 30; DB 21; Length 76;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPDPOG 6
11111
Db 34 PTPDPOG 39

RESULT 15
US-60-169-840-6635
Sequence 6635, Application US/60169840
GENERAL INFORMATION:
APPLICANT: Bonazzi, Vivien
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: CLO00164
CURRENT APPLICATION NUMBER: US/60/169,840
CURRENT FILING DATE: 1999-12-09
NUMBER OF SEQ ID NOS: 9628
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6635
LENGTH: 76
TYPE: PRT
ORGANISM: Human
US-60-169-840-6635

Query Match 85.7%; Score 30; DB 21; Length 76;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPDPOG 6
11111
Db 34 PTPDPOG 39

Search completed: November 16, 2000, 15:37:46
Job time: 96 sec

Thu Nov 16 15:44:23 2000

us-09-308-237b-1.rap

Page 6

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 16, 2000, 15:36:10 ; Search time 13.46 Seconds
(without alignments)
28,268 Million cell updates/sec

Title: US-09-308-237b-1
Sequence: 1 PTPGPG 6

Scoring table: BLOSUM62
Gapop: 10.0, Gapext: 0.5

Searched: 182106 seqs, 63460219 residues
Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100	573	JC4335	anti-mullerian hor
2	32	91.4	792	H81693	outer membrane pro
3	32	91.4	792	B71539	probable omp85 ana
4	31	88.6	374	A42270	alpha (1,3) fucosyl
5	31	88.6	557	S41627	probable anti-mull
6	31	88.6	568	JC5629	mullerian-inhibiti
7	31	88.6	645	T27186	hypothetical prote
8	31	88.6	780	T20366	hypothetical prote
9	30	85.7	298	C75430	uricase - Delnoco
10	30	85.7	573	A33533	cell surface glyco
11	30	85.7	899	S12319	pre-mRNA splicing
12	30	85.7	1160	T27487	hypothetical prote
13	30	85.7	1160	C47648	hypothetical prote
14	30	85.7	1418	D75281	perlecan homolog u
15	30	85.7	2262	T30890	ribonucleoside-dip
16	30	85.7	2277	B47648	calcium channel al
17	30	85.7	2295	T19820	hypothetical prote
18	30	85.7	2482	T19819	hypothetical prote
19	30	85.7	2482	A47648	perlecan homolog u
20	30	85.7	3375	T19821	hypothetical prote
21	30	85.7	4660	T42737	gp330 protein prec
22	29	82.9	40	A42087	lipopeptide matng
23	29	82.9	99	E69267	hypothetical transhyr
24	29	82.9	124	T41250	antigen 5.1 precu
25	29	82.9	162	YA051	antigen 5.1 precu
26	29	82.9	162	A26769	blood-stage antige
27	29	82.9	162	A26780	single-stranded DN
28	29	82.9	166	T14869	hypothetical prote
29	29	82.9	166	T29641	hypothetical prote

30	29	82.9	176	JN0632	single-stranded DN
31	29	82.9	178	DDEC	single-stranded DN
32	29	82.9	182	T41204	A-minor fibrillar p
33	29	82.9	197	T36584	hypothetical prote
34	29	82.9	205	E82032	thiamin phosphate
35	29	82.9	205	A81011	thiamin phosphate
36	29	82.9	212	S33324	carcinoembryonic a
37	29	82.9	231	T27396	hypothetical prote
38	29	82.9	232	T46679	alpha-smooth muscl
39	29	82.9	281	JC4295	heat-shock protein
40	29	82.9	304	T18345	MGC2 protein precu
41	29	82.9	328	E64020	hypothetical prote
42	29	82.9	361	B59099	hypothetical prote
43	29	82.9	370	S55333	protein kinase pim
44	29	82.9	377	T47791	hypothetical prote
45	29	82.9	390	B69370	branched-chain am

ALIGNMENTS

RESULT 1
JC4335
anti-mullerian hormone type II receptor precursor - human
C:Species: Homo sapiens (man)
C:Date: 06-Dec-1995 #sequence_revision 08-Feb-1996 #text_change 04-Mar-2000
C:Accession: JC4335
R:Visser, J.A.; Meluskey, A.; van Beers, T.; Weghuis, D.O.; van Kessel, A.G.; Grootege
Biochem. Biophys. Res. Commun. 215 1029-1036, 1995
A>Title: Structure and chromosomal localization of the human anti-mullerian hormone
A:Reference number: JC4335; MUID:96028015
A:Accession: JC4335
A:Molecule type: mRNA
A:Residues: 1-573 <VIS>
A:Cross-references: GB:X91156; NID:91107671; PID:CAA62593.1; PID:e186046; PID:q11076
C:Comment: This is a receptor for anti-mullerian hormone (see PIR:WFI00X). It plays a
C:Gene: GDB:AMHR2
A:Gene: GDB:AMHR2
A:Cross-references: GDB:696210; OMIM:600956
A:Map position: 12q13-12q13
A:Introns: 16/3; 77/3; 141/3; 167/3; 207/2; 284/2; 322/3; 380/2; 429/3; 475/2
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; hormone receptor; transmembrane protein
F:1-16/Domain: signal sequence #status predicted <Sig>
F:17-573/Product: anti-mullerian hormone type II receptor #status predicted <MAT>
F:17-141/Domain: extracellular hormone binding #status predicted <ELB>
F:142-167/Domain: transmembrane #status predicted <TM>
F:201-512/Domain: protein kinase homology <KIN>

Query Match 100.0%; Score 35; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. NO. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPGPG 6
DB 369 PTPGPG 374

RESULT 2
H81693
outer membrane protein, probable TC0512 [imported] - Chlamydia muridarum (strain N199
C:Species: Chlamydia muridarum, Chlamydia trachomatis Moyn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: H81693
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwalt, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis Moyn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150235
A:Accession: H81693
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-792 <TEP>
 A:Cross-references: GB:AE002320; GB:AE002160; NID:g7190551; PIDN:AA939354.1; PID:g719055
 A:Experimental source: strain N199 (Mopn)
 C:GeneID: 82826
 A:Gene: TC0512

Query Match
 Best Local Similarity 91.4%; Score 32; DB 2; Length 792;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
 DB 692 PROPOG 697

RESULT 3
 B71539
 Probable omp85 analog - Chlamydia trachomatis (serotype D, strain UW3/CX)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: B71539
 R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, A.; Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A:Reference number: A71570; MUID:99000809
 A:Accession: B71539
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-792 <ARN>
 A:Cross-references: GB:AE001297; GB:AE001273; NID:g3328646; PIDN:AAC67834.1; PID:g3328646
 A:Experimental source: serotype D, strain UW-3/CX
 C:GeneID: 82826
 A:Gene: yaeT

Query Match
 Best Local Similarity 91.4%; Score 32; DB 2; Length 792;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
 DB 692 PROPOG 697

RESULT 4
 A42270
 alpha (1,3) fucosyltransferase FUT5 - human
 N:Altornate names: fucosyltransferase 5
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 29-Sep-1999
 C:Accession: A42270; I39046; I39047
 R:Wesdon, B.W.; Maier, R.P.; Larsen, R.D.; Lowe, J.B.
 J. Biol. Chem. 267, 4152-4160, 1992
 A:Title: Isolation of a novel human alpha (1,3)fucosyltransferase gene and molecular clones encoding enzymes with distinct acceptor substrate specificities.
 A:Reference number: A42270; MUID:92156161
 A:Accession: A42270
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-374 <MRS>
 A:Cross-references: GB:M81485; NID:g182490; PIDN:AA98117.1; PID:g1280209
 R:Camerton, H.S.; Szczepaniak, D.; Weston, B.W.
 J. Biol. Chem. 270, 20112-20122, 1995
 A:Title: Expression of human chromosome 19p alpha(1,3)-fucosyltransferase genes in normal cells
 A:Reference number: I39043; MUID:95378269
 A:Accession: I39046
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-374 <MRS>
 A:Cross-references: EMBL:U27329; NID:g967194; PIDN:AAC50188.1; PID:g967195
 A:Accession: I39047

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-374 <RE2>
 A:Cross-references: EMBL:U27330; NID:g967196; PIDN:AAC50189.1; PID:g967197
 C:GeneID: 82826
 A:Gene: TC0512

Query Match
 Best Local Similarity 88.6%; Score 31; DB 2; Length 374;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
 DB 137 PROPOG 142

RESULT 5
 S41627
 Probable anti-mullerian hormone receptor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
 C:Accession: S41627
 R:Baarends, W.M.; van Helmond, M.J.L.; Post, M.; van der Schoot, P.J.C.M.; Hoogerbrug
 Development 120, 189-197, 1994
 A:Title: A novel member of the transmembrane serine/threonine kinase receptor family
 A:Reference number: S41627; MUID:94163972
 A:Accession: S41627
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-557 <BAA>
 A:Cross-references: EMBL:X71916; NID:g453177; PIDN:CAA50731.1; PID:g453178
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 F:199-510/Domain: protein kinase homology <KIN>

Query Match
 Best Local Similarity 88.6%; Score 31; DB 2; Length 557;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
 DB 367 PROPOG 372

RESULT 6
 JC5629
 mullerian-inhibiting substance type II receptor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 04-Mar-2000
 C:Accession: JC5629
 R:Mishina, Y.; Tizard, R.; Deng, J.M.; Pathak, B.G.; Copeland, N.G.; Jenkins, N.A.; C
 Biochem. Biophys. Res. Commun. 237, 741-746, 1997
 A:Title: Sequence, genomic organization, and chromosomal location of the mouse Muc11e
 A:Reference number: JC5629; MUID:97445109
 A:Accession: JC5629
 A:Molecule type: DNA
 A:Residues: 1-568 <MIS>
 C:Comment: This receptor is involved in the sexual differentiation.
 C:GeneID: 82826
 A:Gene: Amhr
 A:Map position: 15
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 F:197-508/Domain: protein kinase homology <KIN>

Query Match
 Best Local Similarity 88.6%; Score 31; DB 2; Length 568;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPPOG 6
111111
Db 365 PTPPOG 370

RESULT 7

T27186
hypothetical protein Y54G9A.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27186

R:Smyle, R.
submitted to the EMBL data library, October 1998

A:Reference number: Z20324

A:Accession: T27186

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-645 <NT>

A:Cross-references: EMBL:AI032648; PIDN:CA21699.1; GSPDB:GN00020; CESP:Y54G9A.3

A:Experimental source: clone Y54G9A

C:Genetics:

A:Gene: CESP:Y54G9A.3

A:Map position: 2

A:Introns: 56/3; 100/2; 148/1; 411/2; 541/2; 575/3

Query Match 88.6%; Score 31; DB 2; Length 645;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPPOG 6
111111
Db 472 PTPPOG 477

RESULT 8

T00366
hypothetical protein KIAA0669 - human

C:Species: *Homo sapiens* (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Oct-1999

C:Accession: T00366

R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.

DNA Res. 5, 169-176, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete

A:Reference number: Z14142; MUID:98403880

A:Accession: T00366

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-780 <1SH>

A:Cross-references: EMBL:AB014569; NID:d1204324; PIDN:BA031644.1; PID:d1032605

A:Experimental source: brain; clone HK02346

C:Genetics:

A:Note: KIAA0669

Query Match 88.6%; Score 31; DB 2; Length 780;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPPOG 6
111111
Db 311 PTPPOG 316

RESULT 9

C75430
uricase - *Deinococcus radiodurans* (strain R1)

C:Species: *Deinococcus radiodurans*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: C75430

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zaleski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250; MUID:20036896

A:Accession: C75430

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <WHT>

A:Cross-references: GB:AE001965; GB:AE000513; NID:96458892; PIDN:AAFI0733.1; PID:9645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1160

A:Map position: 1

C:Superfamily: urate oxidase

Query Match 85.7%; Score 30; DB 2; Length 298;

Best Local Similarity 83.3%; Pred. No. 91;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTPPOG 6
111111
Db 125 PTPPOG 130

RESULT 10

A33533
cell surface glycoprotein precursor - mouse

C:Species: *Mus musculus* (house mouse)

C:Date: 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change 11-Jan-2000

C:Accession: A33533

R:Dougherty, G.J.; Kay, R.J.; Humphries, R.K.

J. Biol. Chem. 264, 6509-6514, 1989

A:Title: Molecular cloning of 114A10, a cell surface antigen containing highly conse

lines.

A:Reference number: A33533; MUID:89197960

A:Accession: A33533

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-573 <DOU>

A:Cross-references: GB:J04634; NID:q191943; PIDN:AAA37239.1; PID:g309106

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: glycoprotein; transmembrane protein

F:429-466/Domain: EGF homology <EGF>

Query Match 85.7%; Score 30; DB 2; Length 573;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTPPOG 6
111111
Db 213 PTPPOG 218

RESULT 11

S12319
pre-mRNA splicing factor PRP6 - yeast (*Saccharomyces cerevisiae*)

N:Allerente names: protein YBR0508; protein YBR050C

C:Species: *Saccharomyces cerevisiae*

C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 17-Mar-1999

C:Accession: S12319; S45913; S49510; S55854

R:Legrain, P.; Choulika, A.

EMBO J. 9, 2775-2781, 1990

A:Title: The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cy

A:Reference number: S12319; MUID:90360988

A:Accession: S12319

A:Molecule type: DNA

A:Residues: 1-899 <LEG>

A:Cross-references: EMBL:X53465; NID:94238; PID:94239

R:Aljunovic, G.; Pohl, F.M.; Pohl, T.M.

submitted to the Protein Sequence Database, August 1994

A:Reference number: 545906
 A:Accession: 545913
 A:Molecule type: DNA
 A:Residues: 1-899 <ALJ>
 A:Cross-references: EMBL:235924; NID:9536250; PID:9536291; MIPS:YBR055C
 A:Experimental source: strain S286C
 R:Aljnovic, G.
 submitted to the EMBL Data Library, October 1994
 A:Description: Sequence and analysis of 24 kb on chromosome II of *Saccharomyces cerevisiae*
 A:Reference number: 549503
 A:Accession: 549510
 A:Molecule type: DNA
 A:Residues: 1-899 <ALJ>
 A:Cross-references: EMBL:246260; NID:9559942; PID:9559951
 A:Experimental source: strain S286C
 R:Aljnovic, G.; Poll, T.M.
 Yeast 11, 475-479, 1995
 A:Title: Sequence and analysis of 24 kb on chromosome II of *Saccharomyces cerevisiae*
 A:Reference number: 555854
 A:Accession: 555854
 A:Molecule type: DNA
 A:Residues: 1-899 <ALW>
 A:Cross-references: EMBL:246260; NID:9559942; PID:9559951
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1994
 A:Gene: SCD:PRO6
 A:Cross-references: SCD:S0000259; MIPS:YBR055C
 A:Map position: 2R
 C:Keywords: nucleus; RNA binding; zinc finger

Query Match
 Best Local Similarity 85.7%; Score 30; DB 2; Length 899;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PTPPOG 6
 Db 238 PTPPOG 243

RESULT 12
 127487
 Hypothetical protein ZC101.2b - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T27487
 R:Percy, C.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: 220375
 A:Accession: T27487
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1160 <WIL>
 A:Cross-references: EMBL:293395; PIDN:CAE07704.1; GSPDB:GN00020; CESP:ZC101.2b
 A:Experimental source: clone ZC101
 C:Genetics:
 A:Gene: CESP:ZC101.2b
 A:Map position: 2
 A:Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2

Query Match
 Best Local Similarity 85.7%; Score 30; DB 2; Length 1160;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PTPPOG 6
 Db 712 PTPPOG 717

RESULT 13
 C47648

perlecan homolog unc-52, form 3 precursor - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 16-Jul-1999
 C:Accession: C47648
 R:Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
 Gene Dev. 7, 1471-1484, 1993
 A:Title: Products of the unc-52 gene in *Caenorhabditis elegans* are homologous to the
 A:Reference number: A47648; MUID:93339574
 A:Accession: C47648
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1160 <ROG>
 A:Cross-references: GB:L13458
 C:Superfamily: LDL receptor ligand-binding repeat homology; laminin-type EGF-like hom
 F:149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:935-1002/Domain: laminin-type EGF-like homology <LEG1>
 F:1011-1058/Domain: laminin-type EGF-like homology <LEG2>

Query Match
 Best Local Similarity 85.7%; Score 30; DB 2; Length 1160;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PTPPOG 6
 Db 712 PTPPOG 717

RESULT 14
 D75281
 ribonucleoside-diphosphate reductase-related protein - *Deinococcus radiodurans* (strat
 C:Species: *Deinococcus radiodurans*
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: D75281
 R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: D75281
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1418 <MHI>
 A:Cross-references: GB:AE002068; GB:AE000513; NID:96460186; PIDN:AAFI1919.1; PID:9646
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2374
 A:Map position: 1

Query Match
 Best Local Similarity 85.7%; Score 30; DB 2; Length 1418;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PTPPOG 6
 Db 233 PTPPOG 238

RESULT 15
 T30890
 calcium channel alpha1-chain - sea squirt (*Halocynthia roretzi*)
 N:Alternate names: ascidian calcium channel alpha1-subunit
 C:Species: *Halocynthia roretzi*
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T30890
 R:Okamura, Y.; Okagaki, R.
 submitted to the EMBL Data Library, May 1998
 A:Description: Tupal, ascidian calcium channel alpha1-subunit.
 A:Reference number: 220924

Thu Nov 16 15:44:23 2000

us-09-308-237b-1.rpr

Page 5

A:Accession: T30890
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2262 <OK>
A:Cross-references: EMBL:AB013604; PIDN:BAA34927.1
A:Experimental source: young tadpole larvae
C:Genetics:
A:Gene: TUCa1
C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain
C:Keywords: calcium binding; calcium channel

Query Match 85.7%; Score 30; DB 2; Length 2262;
Best local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PTQPOG 6
Db 1839 PTQPOG 1844

Search completed: November 16, 2000, 15:37:05
Job time: 55 sec

1

2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 16, 2000, 15:36:34 ; Search time 8.75 Seconds
(without alignments)

21.907 Million cell updates/sec

Title: US-05-308-237b-1
Perfect score: 35
Sequence: 1 PRTPOG 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	573	1 AMH2_HUMAN	Q16671 homo sapien
2	31	88.6	374	1 FUT5_HUMAN	O11128 homo sapien
3	31	88.6	374	1 FUT5_PANTR	P56433 pan troglod
4	31	88.6	557	1 AMH2_RAT	O62893 rattus norv
5	30	85.7	398	1 TRX1_HUMAN	O43435 homo sapien
6	30	85.7	545	1 TRX4_CABEL	O09392 caenorhabdi
7	30	85.7	573	1 C114_MOUSE	P19467 mus musculu
8	30	85.7	899	1 PR06_YEAST	P19735 saccharomyc
9	30	85.7	1394	1 CNG4_BOVIN	O28181 bos taurus
10	30	85.7	2481	1 UN52_CABEL	O06561 caenorhabdi
11	30	85.7	4660	1 LRP2_RAT	P98158 rattus norv
12	29	82.9	40	1 MFAL_USTMA	P31966 ussiliado ma
13	29	82.9	162	1 CRA_PLAFA	P04922 plasmodium
14	29	82.9	175	1 EXP1_PLAFA	P04926 plasmodium
15	29	82.9	175	1 SSB_SERMA	P25766 seriatia ma
16	29	82.9	177	1 HS32_FCOLI	P02339 escherichia
17	29	82.9	281	1 HS32_DICDI	P54658 dictyostella
18	29	82.9	304	1 P300_HUMAN	O43378 mycoplasma
19	29	82.9	328	1 Z1PA_HAEIN	P14313 haemophilus
20	29	82.9	516	1 GUX1_PACH	P13860 phanerochaete
21	29	82.9	568	1 YG54_YEAST	P53109 saccharomyc
22	29	82.9	578	1 YWR1_CABEL	O20471 caenorhabdi
23	29	82.9	657	1 KNOB_PLAFA	P06719 plasmodium
24	29	82.9	708	1 GPF_DICDI	P36417 dictyostella
25	29	82.9	763	1 EXM1_HUMAN	O08050 h. forthead
26	29	82.9	1174	1 KPC1_COCHOE	O42632 cochliobdium
27	29	82.9	1189	1 YH6_YEAST	P47035 saccharomyc
28	29	82.9	1460	1 CAL1_CANFA	O95937 canis famli
29	29	82.9	2414	1 P300_HUMAN	O09472 homo sapien
30	29	82.9	2437	1 NOTC1_BRACE	P46530 brachydanio
31	29	82.9	2444	1 NOTC1_HUMAN	P46531 homo sapien
32	29	82.9	2524	1 NOTC1_XENLA	P21783 xenopus lae
33	29	82.9	2531	1 NOTC1_MOUSE	O01705 mus musculu

34	29	82.9	2531	1 NOTC1_RAT	O07008 rattus norv
35	28	80.0	143	1 TS22_MOUSE	O00992 mus musculu
36	28	80.0	144	1 TS22_HUMAN	O15714 homo sapien
37	28	80.0	246	1 YP73_MYCTU	O50648 mycobacteri
38	28	80.0	308	1 Y775_METTH	O26869 methanobact
39	28	80.0	359	1 OMPA_SERMA	P04845 serratia ma
40	28	80.0	366	1 CAS4_EPHMU	P18503 ephydratia m
41	28	80.0	463	1 D5DR_FUGRU	P53454 tugu rudrip
42	28	80.0	469	1 COG1_BOVIN	P28053 bos taurus
43	28	80.0	582	1 SYN2_HUMAN	O92777 homo sapien
44	28	80.0	586	1 SYN2_RAT	O63537 rattus norv
45	28	80.0	657	1 HOTL_HUMAN	P42357 homo sapien

ALIGNMENTS

RESULT 1
ID AMH2_HUMAN STANDARD: PRT: 573 AA.
AC Q16671; Q13762;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANTI-MULLERIAN HORMONE TYPE II RECEPTOR PRECURSOR (BC 2.7.1.-) (AMH
DE TYPE II RECEPTOR) (MIS TYPE II RECEPTOR) (MIRII).
GN AMH2 OR AMHR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC [1]
RP MEDLINE: 96083584.
RX Imband S., Faure E., Lamarre I., Mattei M.-G., di Clemente N.,
RA Tizard R., Carre-Busebe D., Bevilille C., Tragelhon L., Tonkin C.,
RA Nelson J., McCalliffe M., Biddart J.-W., Lababidi A., Jossio N.,
RA Cate R.L., Picard J.-Y.;
RT "Insensitivity to anti-mullerian hormone due to a mutation in the
RT human anti-mullerian hormone receptor.";
RL Nat. Genet. 11:382-388(1995).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96028015.
RA Visser J.A., McLuskey A., van Beers T., Weghuis D.O., van Kessel A.G.,
RA Grootegoed J.A., Themmen A.P.N.;
RT "Structure and chromosomal localization of the human anti-mullerian
RT hormone type II receptor gene.";
RL Biochem. Biophys. Res. Commun. 215:1029-1036(1995).
[3]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS: Macaulhlin D.T., Maheswaran S., Teixeira J.M.,
RA Masiakos P.T., Khan P.C., Kehas D.J., Kennally M.K., Dombkowski D.M.,
RA Fuller A.F., Shan P.C., Donahoe P.K.;
RT "Human ovarian cancer, cell lines and primary ascites cells, express
RT the human MIS type II, bind, and are responsive to mis.";
RL Clin. Cancer Res. 0:0-0(1999).
CC -1- FUNCTION: RECEPTOR FOR ANTI-MULLERIAN HORMONE.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- DISPESE: DEFECTS IN AMHR2 ARE THE CAUSE OF MALE
CC PSEUDOHERPAPHRODITISM CHARACTERIZED BY A FAILURE OF MULLERIAN
CC DUCT REGRESSION IN OTHERWISE NORMAL MALES.
CC DUCT REGRESSION BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TSPB RECEPTOR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

Query Match	100.0%	Score 35;	DB 1;	Length 573;
Best Local Similarity	100.0%;	Pred. No. 9, 6;		
Matches	6;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
Qy	1 PROPOG 6			
Db	369 PROPOG 374			

RESULT	2
FUT5_HUMAN	
ID	FUT5_HUMAN
AC	011128;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	ALPHA-(1,3)-FUCOSYLTRANSFERASE (EC 2.4.1.65) (GALACTOSIDE 3-L-FUT5).
GN	FUT5.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-PERIPHERAL BLOOD LEUKOCYTES;
RX	MEDLINE; 92156161.
RA	Weston B.W., Nair R.P., Larsen R.D., Lowe J.B.;
RT	"Isolation of a novel human alpha (1,3)fucosyltransferase gene and
RT	molecular comparison to the human Lewis blood group alpha
RT	(1,3)/,4)fucosyltransferase gene. Syntenic, homologous, nonallelic
RT	genes encoding enzymes with distinct acceptor substrate
RT	specificities.";
RL	J. Biol. Chem. 267:4152-4160(1992).
RN	[2]

RP SEQUENCE FROM N.A.
R TISSUE-COLON, KIDNEY, AND LIVER;
RX MEDLINE; 95378269.
RA Cameron H.S., Szczepaniak D., Weston B.W.;
RT "Expression of human chromosome 10p alpha(1,3)-fucosyltransferase
RT genes in normal tissues. Alternative splicing, polyadenylation, and
RT isoforms." ;
RU J. Biol. Chem. 270:20112-20122(1995)
CC -I- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
CC THE EXPRESSION OF VIM-2, LEWIS X/SSA-1 AND STALL LEWIS X
CC ANTIGENS.
CC -I- CATALYTIC ACTIVITY: GDP-L-FUCOSE + 1,3-BETA-D-GALACTOSYL-
CC N-ACETYL-D-GLUCOSAMINYL-R = GDP + 1,3-BETA-D-GALACTOSYL-
CC [ALPHA-1,4-L-FUCOSYL]-N-ACETYL-D-GLUCOSAMINYL-R.
CC -I- PATHWAY: GLYCOSYLATION.
CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC FORM IN TRANS CISTERNAE OF GOLGI.
CC -I- TISSUE SPECIFICITY: LIVER, COLON AND TESTIS AND TRACE AMOUNTS IN
CC T-CELLS AND BRAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81485; AAA98117.1; -
DR EMBL; U27329; AAC50188.1; -
DR EMBL; U27330; AAC50189.1; -
DR MIM; 136835; -
DR INTERPRO: IPR001503; -
DR PFM; PF00852; Fucosyl1.transf.1;
KW Signal-anchor: Glycosyltransferase; Glycoprotein; Transmembrane;
KM Signal-anchor: Golgi stack.
FT DOMAIN 1 15
FT DOMAIN 16 34
FT FT
FT FT
FT DOMAIN
FT FT
FT CARBOHYD 35 374
FT CARBOHYD 60 60
FT CARBOHYD 105 105
FT CARBOHYD 167 167
FT CARBOHYD 198 198
SQ SEQUENCE 374 AA; 43008 MW; BB25B81523... (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
LUMENAL, CATALYTIC (POTENTIAL).
N-LINKED (GLCNCG . . .) (POTENTIAL).
N-LINKED (GLCNCG . . .) (POTENTIAL).
N-LINKED (GLCNCG . . .) (POTENTIAL).
N-LINKED (GLCNCG . . .) (POTENTIAL).

RA Henry S., Blanchet A., Iordachescu D., Oriol R., Mollicone R.;
 RT "Evolution of fucosyltransferase genes in vertebrates.";
 RL J. Biol. Chem. 272:29721-29728(1997)
 CC -1- FUNCTION: MAY CATALYZE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
 CC THE EXPRESSION OF YIM-2, LEWIS X/SSEA-1 AND SIALYL LEWIS X
 CC ANTIGENS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: GDP-L-FUCOSE + 1,3-BETA-D-GALACTOSYL-
 CC N-ACETYL-D-GLUCOSAMINYL-R = GDP + 1,3-BETA-D-GALACTOSYL-
 CC (ALPHA-1,4-L-FUCOSYL)-N-ACETYL-D-GLUCOSAMINYL-R.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 CC FORM IN TRANS CISTERNAE OF GOLGI.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y14034; CA574361.1;
 CC DR INTERPRO: IPR001503;
 CC DR PFAM: PF00852; Fucosyl_1transf; 1.
 CC KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
 CC Signal-anchor; Golgi stack.
 CC FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 16 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC FT (POTENTIAL).
 CC FT DOMAIN 35 374 IUDENAL, CATALYTIC (POTENTIAL).
 CC FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 374 AA; 43034 MW; 3F35DEB8A8F1A222 CRC64;
 CC SEQUENCE

Query Match 88.6%; Score 31; DB 1; Length 374;
 Best Local Similarity 83.3%; Pred. No. 38;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPOG 6
 Db 137 PTPOG 142
 11:1111

RESULT 4
 ID AMH2.RAT STANDARD; PRT; 557 AA.
 AC 062893; 063045; 09ROA7;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ANTI-MULLERIAN HORMONE TYPE II RECEPTOR PRECURSOR (EC 2.7.1.-) (AMH
 DE TYPE II RECEPTOR) (MIS TYPE II RECEPTOR) (MIRIT) (C14).
 GN AMH2.
 OS Rattus norvegicus (Rat).
 OS Eukaryota; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 96107132.
 RA Teixeira J., Ha W.M., Shah P.C., Morikawa N., Lee M.M., Catlin E.A.,
 RA Hudson P.L., Wing J., MacLaughlin D.T., Donahoe P.K.;
 RT "Developmental expression of a candidate müllerian inhibiting
 RT substance type II receptor.";
 RL Endocrinology 137:160-165(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RX MEDLINE: 94163972.
 RX Baarends W.M., Van Helmond M.J.L., Post M., Van der Schoot P.J.C.M.,
 RA Hoogerbrugge J.W., de Winter J.P., Uilenbroek J.T.J., Karels B.,

RA Wilming L.G., Meijers J.H.C., Themmen A.P.N., Grootegeed A.J.;
 RT "A novel member of the transmembrane serine/threonine kinase receptor
 RT family is specifically expressed in the gonads and in mesenchymal
 RT cells adjacent to the müllerian duct.";
 RL Development 120:189-197(1994).
 CC [3]
 RN SEQUENCE OF 1-16 FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RA Teixeira J., Keras D.J., Antun R., Donahoe P.K.;
 RT "Transcriptional regulation of the rat müllerian inhibiting substance
 RT type II receptor in rodent Leydig cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13831-13838(1999).
 CC -1- FUNCTION: RECEPTOR FOR ANTI-MULLERIAN HORMONE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE MESENCHYMAL CELLS
 CC SURROUNDING THE MULLERIAN DUCT AT EMBRYONIC DAYS 14, 15, AND 16
 CC AND IN TUBULAR AND FOLLICULAR STRUCTURES OF THE FETAL GONADS.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC TGBB RECEPTOR SUPERFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U42427; AAC52343.1;
 CC DR EMBL: X71916; CA50731.1;
 CC DR EMBL: AF092445; AAC64138.1;
 CC DR INTERPRO: IPR000472;
 CC DR INTERPRO: IPR000719;
 CC DR PFAM: PF00069; pkinase; 1.
 CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
 CC DR PROSITE: PS00108; PROTEIN_KINASE_ST_FALSE_NEG.
 CC DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 CC KW Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding;
 CC Transmembrane; Glycoprotein; Signal.
 CC FT SIGNAL 1 17 POTENTIAL.
 CC FT CHAIN 18 57 ANTI-MULLERIAN HORMONE TYPE II RECEPTOR.
 CC FT DOMAIN 18 144 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 145 165 POTENTIAL.
 CC FT DOMAIN 166 557 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 201 511 PROTEIN KINASE.
 CC FT NR BIND 207 215 ATP (BY SIMILARITY).
 CC FT BINDING 228 228 ATP (BY SIMILARITY).
 CC FT ACT_SITE 331 331 BY SIMILARITY.
 CC FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CONFLICT 527 527 C->Y (IN REF. 2).
 CC FT SEQUENCE 557 AA; 8EDBE9C0C32BEDD5 CRC64;
 CC SEQUENCE

Query Match 88.6%; Score 31; DB 1; Length 557;
 Best Local Similarity 83.3%; Pred. No. 58;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPOG 6
 Db 367 PTPOG 372
 11:1111

RESULT 5
 ID TBX1.HUMAN STANDARD; PRT; 398 AA.
 AC 043435; 043436;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TBX1 PROTEIN (T-BOX PROTEIN 1) (TESTIS-SPECIFIC T-BOX PROTEIN).
 GN TBX1.
 OS Homo sapiens (Human).

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SOURCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE, AND TESTIS;
RX MEDLINE: 97422603.
RA Chieffo C., Garvey N., Gong W., Roe B., Zhang G., Silver L.,
RA Emanuel B.S., Budarf M.L.;
RT "Isolation and characterization of a gene from the Digeorge
RT chromosomal region homologous to the mouse Tbx1 gene."
RL Genomics 43:267-277(1997)
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: AF012130; AB94018.1; -
DR EMBL: AF012131; AB94019.1; -
DR MIM: 602054; -
DR INTERPRO: IP001699; -
DR PRAM: PR00907; T-box; 1.
DR PRINTS: PR00937; TBOX.
DR PROSITE: PS01283; TBOX_1; 1.
DR PROSITE: PS01264; TBOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 43 48 POLY-PRO.
FT DOMAIN 54 57 POLY-ALA.
FT DOMAIN 61 67 POLY-PRO.
FT DOMAIN 94 99 POLY-ALA.
FT DNA_BIND 119 297 T-BOX.
FT VASPLIC 338 398 GHVLRKDKYKAKTSRNPPEVEVLLRDAGCCVNLGCPAE
FT KPSKKSSRLSPKPKDT (IN ISOFORM B).
SQ SEQUENCE 398 AA: 43133 MW: 84033.00 CRC64;
-----
Query Match 85.7%; Score 30; DB 1; Length 398;
Best Local Similarity 85.3%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PROPOG 6
DB 328 PROPOG 333
-----
RESULT 6
YR49_CAEEL STANDARD; PRT; 545 AA.
AC 009382;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHEICAL 61.9 KDA PROTEIN F47D12.9 IN CHROMOSOME III.
GN F47D12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
RN Rhabditidae; Peloderinae; Caenorhabditis.
RP [1]
RC SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RA Tatch A.;
RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: U22831; AAA64320.1; -
DR WORKREP: F47D12.9; CE01952.
KW Hypothetical protein.
SQ SEQUENCE 545 AA: 61946 MW: 50F0BD70CE9A343C CRC64;
-----
Query Match 85.7%; Score 30; DB 1; Length 545;
Best Local Similarity 83.3%; Pred. No. 89;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PROPOG 6
DB 77 PROPOG 82
-----
RESULT 7
C14_MOUSE STANDARD; PRT; 573 AA.
ID C14_MOUSE
AC P19467;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CELL SURFACE ANTIGEN 114/A10 PRECURSOR.
GN Lf64.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89197960.
RA Dougherty G.J., Kay R.J., Humphries R.K.;
RT "Molecular cloning of 114/A10, a cell surface antigen containing
RT highly conserved repeated elements, which is expressed by murine
RT hemopoietic progenitor cells and interleukin-3-dependent cell
RT lines."
RL J. Biol. Chem. 264:6509-6514(1989).
CC -1- FUNCTION: MAY HAVE A POSITIVE REGULATORY ROLE IN THE CELLULAR
CC RESPONSE TO IL-3.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- PTM: THE SER/THR-RICH TANDM REPEATS MAY SERVE AS SITES OF
CC EXTENSIVE GLYCOSYLATION.
CC -1- THE MOTIF SER-GLY CONSERVED IN ALL 8 SER/THR-RICH REPEATS MAY
CC SERVE AS THE SITE OF ATTACHMENT OF GLYCOSAMINOGLICAN SIDE CHAINS.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: J04634; AAA37239.1; -
DR PIR: A33533; A33533.
DR MGD: MGI:103190; LY64.
DR INTERPRO: IPR000882; -
DR INTERPRO: IPR000561; -
DR PFAM: PF01390; SEA; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 2.
KW Glycoprotein; Signal; Antigen; EGF-like domain; Repeat; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 573 POTENTIAL.
FT DOMAIN 18 480 CELL SURFACE ANTIGEN 114/A10.
FT TRANSMEM 481 508 EXTRACELLULAR (POTENTIAL).

```


FT DOMAIN 509 573 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 17 230 8 X TANDEM REPEATS, SER/THR-RICH.
 FT REPEAT 17 43 1.
 FT REPEAT 44 70 2.
 FT REPEAT 71 97 3.
 FT REPEAT 98 124 4.
 FT REPEAT 125 151 5.
 FT REPEAT 152 178 6.
 FT REPEAT 179 205 7 (NEAR PERFECT).
 FT REPEAT 206 230 8 (APPROXIMATE).
 FT REPEAT 233 273 EGF-LIKE 1.
 FT DOMAIN 385 425 EGF-LIKE 2.
 FT DOMAIN 425 467 EGF-LIKE 3.
 FT DISULFID 237 248 BY SIMILARITY.
 FT DISULFID 242 257 BY SIMILARITY.
 FT DISULFID 259 272 BY SIMILARITY.
 FT DISULFID 389 402 BY SIMILARITY.
 FT DISULFID 394 408 BY SIMILARITY.
 FT DISULFID 410 424 BY SIMILARITY.
 FT DISULFID 429 441 BY SIMILARITY.
 FT DISULFID 433 451 BY SIMILARITY.
 FT DISULFID 453 466 BY SIMILARITY.
 FT CARBOHYD 266 266 N-LINKED (GLCNAC).
 FT CARBOHYD 316 316 N-LINKED (GLCNAC).
 FT CARBOHYD 397 397 N-LINKED (GLCNAC).
 SQ SEQUENCE 573 AA; 58701 MM; 1154CAF04E2D58A9 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 573;
 Best Local Similarity 83.3%; Pred. No. 94;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PROPOG 6
 Db 213 PTPPSG 218

RESULT 8
 PRO6_YEAST STANDARD; PRT; 899 AA.
 ID PRO6_YEAST
 AC P19735.
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, last sequence update)
 DT 01-FEB-1996 (Rel. 33, last annotation update)
 DE PRE-MRNA SPLICING FACTOR PRP6
 DE PRP6 OR RNA6 OR YBR055C OR YBR0508.
 GN Saccharomyces cerevisiae (Baker's yeast).
 OS Saccharomyces cerevisiae
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetes;
 CC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN-S288C / GRF88;
 RC MEDLINE: 90360588.
 RA Legrain P., Chculika A.;
 RT "The molecular characterization of PRP6 and PRP9 yeast genes reveals
 RL a new cysteine/histidine motif common to several splicing factors.";
 RL EMBO J. 9:2775-2781(1990).
 RN [2]
 RP SEQUENCE FROM H.A.
 RP STRAIN-S288C;
 RC MEDLINE: 95321020.
 RA Allinovic G., Pohl T.M.;
 RT "Sequence and analysis of 24 kb on chromosome II of Saccharomyces
 RL cerevisiae.";
 RL Yeast 11:475-479(1995).
 CC -1- FUNCTION: INVOLVED IN PRE-MRNA SPLICING.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- SIMILARITY: TO SOME OTHER PRE-MRNA SPLICING FACTORS.
 CC -1- SIMILARITY: CONTAINS 10 TPR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X53465; CAA37559.1; -
 DR EMBL: 235924; CAA84998.1; -
 DR EMBL: 246260; CAA86398.1; -
 DR FIR: S12319; S12319.
 DR SGD: S0000259; PRP6.
 KW mRNA processing; mRNA splicing; tpr domain; Repeat; Nuclear protein.
 FT REPEAT 225 228 TPR.
 FT REPEAT 259 290 TPR.
 FT REPEAT 291 321 TPR.
 FT REPEAT 322 352 TPR.
 FT REPEAT 353 384 TPR.
 FT REPEAT 385 414 TPR.
 FT REPEAT 497 528 TPR.
 FT REPEAT 686 717 TPR.
 FT REPEAT 755 786 TPR.
 FT REPEAT 855 886 TPR.
 SQ SEQUENCE 899 AA; 104228 MM; 3E87F8AF63EAC41 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 899;
 Best Local Similarity 83.3%; Pred. No. 1,5e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PROPOG 6
 Db 238 PTPPSG 243

RESULT 9
 CNG4_BOVIN STANDARD; PRT; 1394 AA.
 ID CNG4_BOVIN
 AC 028181; 028082; 003861;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 30-MAY-2000 (Rel. 39, last annotation update)
 DE 240 KDA PROTEIN OF ROD PHOTORECEPTOR CNG-CHANNEL [CONTAINS: GLUTAMIC
 DE ACID-RICH PROTEIN (GARP); CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 4
 DE (CNG CHANNEL 4) (CNG-4) (CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
 DE MODULATORY SUBUNIT)].
 GN CNG4.
 OS Bos taurus (Bovine).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Euteleia; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RP MEDLINE: 96009859.
 RA Koerschgen H.G., Illing M., Seifert R., Sefti F., Williams A.,
 RA Gotz U.B., Colville C., Mueller F., Dose A., Godde M., Molday L.,
 RA Kupp U.B., Molday R.S.;
 RT "A 240 kDa protein represents the complete beta subunit of the cyclic
 RL nucleotide-gated channel from rod photoreceptor.";
 RL Neuron 15:627-636(1995).
 RN [2]
 RP SEQUENCE OF 454-1394 FROM N.A.
 RP TISSUE-TESTIS;
 RC MEDLINE: 96198098.
 RA Biel M., Zong X., Ludwig A., Sauter A., Hofmann F.;
 RT "Molecular cloning and expression of the modulatory subunit of the
 RL cyclic nucleotide-gated cation channel.";
 RL J Biol. Chem. 271:6349-6355(1996).
 RN [3]
 RP SEQUENCE OF 1-590 FROM N.A.
 RP TISSUE-RETINA;
 RC Sugimoto Y., Tatsunami K., Tsujimoto M., Khorana H.G., Ichikawa A.;
 RA Sugimoto Y., Tatsunami K., Tsujimoto M., Khorana H.G., Ichikawa A.;
 RL Submitted (XX-1991) to the EMBL/Genbank/DBD databases.
 CC -1- SUBUNIT: FORMS FUNCTIONAL HETEROLOGIC CHANNELS WITH CNG3.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

-1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; CNG4C (SHOWN HERE), CNG4D AND CNG4E: ARE PRODUCED BY ALTERNATIVE SPLICING. CNG4D IS BY FAR THE MOST FREQUENT FORM (CNG4D:CNG4E = 20:1) IN TESTS. -1- TISSUE SPECIFICITY: RETINA, TESTIS, KIDNEY, HEART, AND BRAIN. -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL FAMILY.

 THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/> OR SEND AN EMAIL TO license@isb-sib.ch).

DR EMBL: X89626; CAN61769.1; -
 DR EMBL: X94707; CAN64367.1; -
 DR EMBL: M61185; AAA30536.1; -
 DR INTERPRO: IPR000595; -
 DR INTERPRO: IPR002025; -
 DR PFAM: PF00914; CNG_membrane; 1.
 DR PROSITE: PS00027; CNG_BINDING_1; 1.
 DR PROSITE: PS00888; CNG_BINDING_1; 1.
 DR PROSITE: PS00889; CNG_BINDING_2; 1.
 DR PROSITE: PS50042; CNG_BINDING_3; 1.
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane;
 KW Multigene family; Alternative splicing.
 FT CHAIN 1 590
 FT DOMAIN 1 1394
 FT TRANSMEM 1 767
 FT DOMAIN 1 767
 FT TRANSMEM 768 786
 FT DOMAIN 787 800
 FT TRANSMEM 801 819
 FT DOMAIN 821 844
 FT TRANSMEM 845 864
 FT DOMAIN 865 901
 FT TRANSMEM 902 924
 FT DOMAIN 925 968
 FT TRANSMEM 969 988
 FT DOMAIN 989 1072
 FT TRANSMEM 1073 1093
 FT DOMAIN 1094 1394
 FT NP_BIND 1081 1219
 FT BINDING 1141 1141
 FT CARBOHYD 1153 1153
 FT BINDING 1067 1067
 FT VARSPLIC 515 532
 FT VARSPLIC 532 530
 FT CONFLICT 341 341
 FT CONFLICT 454 465
 FT CONFLICT 482 499
 FT CONFLICT 499 572
 FT CONFLICT 572 590
 FT CONFLICT 1283 1283
 FT CONFLICT 1289 1289
 FT CONFLICT 1336 1336
 FT CONFLICT 1338 1338
 FT SEQUENCE 1394 AA; 155064 MW; EEE6DA559BE3744A7 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 1394;
 Best Local Similarity 83.3%; Pred. No. 2, 3e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PROPOG 6
 DB 655 PROPOG 660

RESULT 10
 UNS2_CABEL
 ID UNS2_CABEL
 STANDARD: PRT; 2481 AA.

AC 006561;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BASEMENT MEMBRANE PROTEOLYCAN PRECURSOR (PERLECAN HOMOLOG).
 GN UNC-52.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 RN Rhabditidae; Peloderae; Caenorhabditis.
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93339574.
 RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;
 RT "Products of the unc-52 gene in Caenorhabditis elegans are homologous
 RT to the core protein of the mammalian basement membrane heparan
 RT sulfate proteoglycan."
 RL Genes Dev. 7:1471-1484 (1993).
 CC -1- FUNCTION: PROBABLE ROLE IN MYOFILAMENT ASSEMBLY AND/OR ATTACHMENT
 CC OF THE MYOFILAMENT LATTICE TO THE CELL MEMBRANE. UNC-52 MAY BE AN
 CC EXTRACELLULAR ANCHOR FOR INTEGRIN RECEPTORS IN MUSCLE.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- ALTERNATIVE PRODUCTS: UNC-52 PRODUCES AT LEAST THREE POLYPEPTIDES:
 CC ONE VERY SHORT FORM AND TWO LONG FORMS.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANE OF ALL
 CC AND M-LINES WHICH ARE ASSOCIATED WITH BETA-INTEGRIN.
 CC -1- DEVELOPMENTAL STAGE: SYNTHESIZED EARLY IN EMBRYOGENESIS.
 CC -1- SIMILARITY: CONTAINS 3 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 16 IMINOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: TO PERLECAN.
 CC -----
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>
 CC OR SEND AN EMAIL TO license@isb-sib.ch).
 CC -----
 DR EMBL: L13458; AAA28156.1; -
 DR HSSP: P01130; 1AUJ.
 DR INTERPRO: IPR000034; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR002049; -
 DR INTERPRO: IPR002172; -
 DR INTERPRO: IPR003006; -
 DR PFAM: PF00047; 1g; 16.
 DR PFAM: PF00052; laminin_B; 2.
 DR PFAM: PF00057; laminin_EGF; 5.
 DR PROSITE: PS00022; EGF_1; 4.
 DR PROSITE: PS00053; laminin_EGF; 3.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01209; LDLR_1; 3.
 DR PROSITE: PS50068; LDLR_2; 3.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 7.
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;
 KW Basement membrane; Extracellular matrix; Alternative splicing;
 KW Laminin EGF-like domain.
 FT SIGNAL 1 18
 FT CHAIN 1 2481
 FT DOMAIN 1 19
 FT DOMAIN 19 148
 FT DOMAIN 149 188
 FT DOMAIN 190 229
 FT DOMAIN 233 272
 FT DOMAIN 273 363
 FT DOMAIN 364 431
 FT DOMAIN 432 441
 FT DOMAIN 442 632
 FT DOMAIN 633 665
 FT DOMAIN 673 719
 FT DOMAIN 720 729
 FT POTENTIAL.
 FT BASEMENT MEMBRANE PROTEOLYCAN.
 FT IG-LIKE C2-TYPE DOMAIN 1.
 FT LDL-RECEPTOR CLASS A 1.
 FT LDL-RECEPTOR CLASS A 2.
 FT LDL-RECEPTOR CLASS A 3.
 FT IG-LIKE C2-TYPE DOMAIN 2.
 FT LAMININ EGF-LIKE 1 (INCOMPLETE).
 FT LAMININ EGF-LIKE 2 (N-TERMINAL).
 FT LAMININ EGF-LIKE 3 (C-TERMINAL).
 FT LAMININ EGF-LIKE 4 (N-TERMINAL).
 FT LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT LAMININ EGF-LIKE 6 (N-TERMINAL).

FT DOMAIN 730 920 LAMININ DOMAIN IV 2.
 FT DOMAIN 921 953 LAMININ EGF-LIKE 4 (C-TERMINAL).
 FT DOMAIN 954 1003 LAMININ EGF-LIKE 5.
 FT DOMAIN 1010 1059 LAMININ EGF-LIKE 6.
 FT DOMAIN 1060 1110 LAMININ EGF-LIKE 7.
 FT DOMAIN 1133 1216 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 1227 1311 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 1315 1403 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 1415 1499 IG-LIKE C2-TYPE DOMAIN 6.
 FT DOMAIN 1507 1592 IG-LIKE C2-TYPE DOMAIN 7.
 FT DOMAIN 1598 1682 IG-LIKE C2-TYPE DOMAIN 8.
 FT DOMAIN 1769 1785 IG-LIKE C2-TYPE DOMAIN 9.
 FT DOMAIN 1794 1878 IG-LIKE C2-TYPE DOMAIN 10.
 FT DOMAIN 1887 1976 IG-LIKE C2-TYPE DOMAIN 11.
 FT DOMAIN 1977 2070 IG-LIKE C2-TYPE DOMAIN 12.
 FT DOMAIN 2078 2162 IG-LIKE C2-TYPE DOMAIN 13.
 FT DOMAIN 2174 2260 IG-LIKE C2-TYPE DOMAIN 14.
 FT DOMAIN 2267 2347 IG-LIKE C2-TYPE DOMAIN 15.
 FT DOMAIN 2354 2436 IG-LIKE C2-TYPE DOMAIN 16.
 FT DISULFID 66 114 BY SIMILARITY.
 FT DISULFID 149 161 BY SIMILARITY.
 FT DISULFID 156 174 BY SIMILARITY.
 FT DISULFID 168 183 BY SIMILARITY.
 FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 197 213 BY SIMILARITY.
 FT DISULFID 209 224 BY SIMILARITY.
 FT DISULFID 233 246 BY SIMILARITY.
 FT DISULFID 240 259 BY SIMILARITY.
 FT DISULFID 253 268 BY SIMILARITY.
 FT DISULFID 954 963 BY SIMILARITY.
 FT DISULFID 956 970 BY SIMILARITY.
 FT DISULFID 973 982 BY SIMILARITY.
 FT DISULFID 985 1001 BY SIMILARITY.
 FT DISULFID 1010 1020 BY SIMILARITY.
 FT DISULFID 1012 1026 BY SIMILARITY.
 FT DISULFID 1029 1038 BY SIMILARITY.
 FT DISULFID 1041 1057 BY SIMILARITY.
 FT DISULFID 1060 1068 BY SIMILARITY.
 FT DISULFID 1062 1078 BY SIMILARITY.
 FT DISULFID 1081 1090 BY SIMILARITY.
 FT DISULFID 1093 1108 BY SIMILARITY.
 FT DISULFID 1151 1199 BY SIMILARITY.
 FT DISULFID 1137 1383 BY SIMILARITY.
 FT DISULFID 1434 1480 BY SIMILARITY.
 FT DISULFID 1526 1572 BY SIMILARITY.
 FT DISULFID 1617 1662 BY SIMILARITY.
 FT DISULFID 1718 1766 BY SIMILARITY.
 FT DISULFID 1813 1860 BY SIMILARITY.
 FT DISULFID 1906 1953 BY SIMILARITY.
 FT DISULFID 1987 2052 BY SIMILARITY.
 FT DISULFID 2098 2146 BY SIMILARITY.
 FT DISULFID 2194 2241 BY SIMILARITY.
 FT DISULFID 2283 2328 BY SIMILARITY.
 FT DISULFID 2373 2419 BY SIMILARITY.
 FT CARBOHYD 1421 1421 N-LINKED (GLCNAC).
 SO SEQUENCE 2481 AA; 270812 MW; E47BEB310F5031 CRC64; (POTENTIAL).

Query Match 85.7%; Score 30; DB 1; Length 2481;
 Best Local Similarity 83.3%; Pred. No. 4.2e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PPOPG 6
 DB 711 PTPNG 716

RESULT 11
 LRP2_RAT STRAND: PPT; 4660 AA.
 AC P98158;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 2 PRECURSOR (MEGALIN)
 DE (GLYCOPROTEIN 330).
 GN LRP2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY;
 RX MEDLINE; 95024033
 RA Saito A., Pletomono S., Loo A.K.C., Fargnhar M.G.;
 RT Complete cloning and sequencing of rat gp330/megalin, a
 RT distinctive member of the low density lipoprotein receptor gene
 RT family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9725-9729(1994).
 RN [2]
 RP FUNCTION.
 RX MEDLINE; 95386696.
 RA Moestrup S.K., Cui S., Vorum H., Bregengaard C., Bjorn S.E.,
 RA Norris K., Gileman J., Christensen E.L.;
 RT Evidence that epithelial glycoprotein 330/megalin mediates uptake of
 RT polybasic drugs.";
 RL J. Clin. Invest. 96:1404-1413(1995).
 RN [3]
 RP TISSUE SPECIFICITY.
 RX MEDLINE; 94172242.
 RA Zheng G., Bachinsky D.R., Stamenkovic I., Strickland D.K., Brown D.,
 RA Andrus G., McCluskey R.T.;
 RT Organ distribution in rats of two members of the low-density
 RT lipoprotein receptor gene family, gp330 and Lrp/alpha 2MR, and the
 RT receptor-associated protein (RAP).";
 RL J. Histochem. Cytochem. 42:531-542(1994).
 CC -1- FUNCTION: BINDS PLASMINOGEN, EXTRACELLULAR MATRIX COMPONENTS,
 CC PLASMINOGEN ACTIVATOR-PLASMINOGEN ACTIVATOR INHIBITOR TYPE 1
 CC COMPLEX, APOLIPOPROTEIN E-ENRICHED BETA-VLDL, LIPOPROTEIN LIPASE,
 CC LACTOFERRIN, CLUSTERIN AND CALCIUM.
 CC -1- FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS
 CC APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B.
 CC -1- SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR-
 CC ASSOCIATED PROTEIN (RAP).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXPRESSED IN
 CC CLEAVAGE AT THE CELL SURFACE.
 CC -1- TISSUE SPECIFICITY: EPITHELIAL CELLS OF KIDNEY GLOMERULUS AND
 CC PROXIMAL TUBULE, LUNG, EPIDIDYMIS, YOLK SAC, AMONG OTHERS.
 CC -1- SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 37 LDL-RECEPTOR CLASS B DOMAINS.
 CC -1- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 SH2-BINDING DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH2-BINDING DOMAIN.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; L34049; AAA51369.1; -.
 CC HSSP; P01130; IATJ.
 DR INTERPRO: IPR000033; -.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR001861; -.
 DR INTERPRO: IPR002172; -.
 DR PFAM; PF00057; 1d1_recept_a; 36.
 DR PFAM; PF00058; 1d1_recept_b; 33.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 8.

DR PROSITE: PS01187; EGF_CA; 3.
 DR PROSITE: PS01209; LDLRA_1; 31.
 DR PROSITE: PS50068; LDLRA_2; 36.
 KW Glycoprotein; Repeat; Endocytosis; Coated pits; Transmembrane;
 KM Receptor; EGF-Like domain; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 4660
 FT DOMAIN 26 4425
 FT TRANSMEM 4426 4446
 FT DOMAIN 4447 4660
 FT DOMAIN 26 64
 FT DOMAIN 65 105
 FT DOMAIN 106 144
 FT DOMAIN 145 181
 FT DOMAIN 181 219
 FT DOMAIN 220 258
 FT DOMAIN 263 307
 FT DOMAIN 308 346
 FT DOMAIN 347 385
 FT DOMAIN 435 476
 FT DOMAIN 478 519
 FT DOMAIN 521 566
 FT DOMAIN 568 611
 FT DOMAIN 612 652
 FT DOMAIN 658 704
 FT DOMAIN 752 793
 FT DOMAIN 795 835
 FT DOMAIN 837 879
 FT DOMAIN 881 923
 FT DOMAIN 969 1013
 FT DOMAIN 1023 1061
 FT DOMAIN 1064 1103
 FT DOMAIN 1108 1146
 FT DOMAIN 1148 1186
 FT DOMAIN 1186 1225
 FT DOMAIN 1229 1269
 FT DOMAIN 1270 1308
 FT DOMAIN 1311 1351
 FT DOMAIN 1351 1390
 FT DOMAIN 1391 1430
 FT DOMAIN 1479 1520
 FT DOMAIN 1522 1563
 FT DOMAIN 1567 1609
 FT DOMAIN 1611 1654
 FT DOMAIN 1656 1696
 FT DOMAIN 1701 1742
 FT DOMAIN 1791 1832
 FT DOMAIN 1834 1882
 FT DOMAIN 1884 1930
 FT DOMAIN 1932 1972
 FT DOMAIN 1973 2013
 FT DOMAIN 2019 2060
 FT DOMAIN 2108 2136
 FT DOMAIN 2158 2201
 FT DOMAIN 2203 2245
 FT DOMAIN 2247 2289
 FT DOMAIN 2291 2332
 FT DOMAIN 2343 2384
 FT DOMAIN 2432 2477
 FT DOMAIN 2479 2518
 FT DOMAIN 2520 2562
 FT DOMAIN 2564 2604
 FT DOMAIN 2605 2647
 FT DOMAIN 2652 2694
 FT DOMAIN 2699 2739
 FT DOMAIN 2740 2778
 FT DOMAIN 2779 2820
 FT DOMAIN 2821 2862
 FT DOMAIN 2863 2903
 FT DOMAIN 2906 2947
 FT DOMAIN 2948 2992
 FT DOMAIN 2993 3031

POTENTIAL.
 POTENTIAL.
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 LDL-RECEPTOR CLASS A 1.
 LDL-RECEPTOR CLASS A 2.
 LDL-RECEPTOR CLASS A 3.
 LDL-RECEPTOR CLASS A 4.
 LDL-RECEPTOR CLASS A 5.
 LDL-RECEPTOR CLASS A 6.
 LDL-RECEPTOR CLASS A 7.
 EGF-LIKE 1.
 EGF-LIKE 2.
 LDL-RECEPTOR CLASS B 1.
 LDL-RECEPTOR CLASS B 2.
 LDL-RECEPTOR CLASS B 3.
 LDL-RECEPTOR CLASS B 4.
 LDL-RECEPTOR CLASS B 5.
 EGF-LIKE 3.
 LDL-RECEPTOR CLASS B 6.
 LDL-RECEPTOR CLASS B 7.
 LDL-RECEPTOR CLASS B 8.
 LDL-RECEPTOR CLASS B 9.
 EGF-LIKE 4.
 LDL-RECEPTOR CLASS A 8.
 LDL-RECEPTOR CLASS A 9.
 LDL-RECEPTOR CLASS A 10.
 LDL-RECEPTOR CLASS A 11.
 LDL-RECEPTOR CLASS A 12.
 LDL-RECEPTOR CLASS A 13.
 LDL-RECEPTOR CLASS A 14.
 LDL-RECEPTOR CLASS A 15.
 EGF-LIKE 5.
 EGF-LIKE 6. CALCIUM-BINDING (POTENTIAL).
 LDL-RECEPTOR CLASS B 10.
 LDL-RECEPTOR CLASS B 11.
 LDL-RECEPTOR CLASS B 12.
 LDL-RECEPTOR CLASS B 13.
 LDL-RECEPTOR CLASS B 14.
 EGF-LIKE 7.
 LDL-RECEPTOR CLASS B 15.
 LDL-RECEPTOR CLASS B 16.
 LDL-RECEPTOR CLASS B 17.
 LDL-RECEPTOR CLASS B 18.
 LDL-RECEPTOR CLASS B 19.
 EGF-LIKE 8.
 LDL-RECEPTOR CLASS B 20.
 LDL-RECEPTOR CLASS B 21.
 LDL-RECEPTOR CLASS B 22.
 LDL-RECEPTOR CLASS B 23.
 LDL-RECEPTOR CLASS B 24.
 EGF-LIKE 9.
 LDL-RECEPTOR CLASS B 25.
 LDL-RECEPTOR CLASS B 26.
 LDL-RECEPTOR CLASS B 27.
 LDL-RECEPTOR CLASS B 28.
 LDL-RECEPTOR CLASS B 29.
 EGF-LIKE 10.
 LDL-RECEPTOR CLASS A 16.
 LDL-RECEPTOR CLASS A 17.
 LDL-RECEPTOR CLASS A 18.
 LDL-RECEPTOR CLASS A 19.
 LDL-RECEPTOR CLASS A 20.
 LDL-RECEPTOR CLASS A 21.
 LDL-RECEPTOR CLASS A 22.
 LDL-RECEPTOR CLASS A 23.

FT DOMAIN 3032 3072 LDL-RECEPTOR CLASS A 24.
 FT DOMAIN 3075 3112 LDL-RECEPTOR CLASS A 25.
 FT DOMAIN 3113 3153 EGF-LIKE 11.
 FT DOMAIN 3154 3194 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 3241 3282 LDL-RECEPTOR CLASS B 30.
 FT DOMAIN 3284 3323 LDL-RECEPTOR CLASS B 31.
 FT DOMAIN 3325 3377 LDL-RECEPTOR CLASS B 32.
 FT DOMAIN 3379 3420 LDL-RECEPTOR CLASS B 33.
 FT DOMAIN 3421 3461 LDL-RECEPTOR CLASS B 34.
 FT DOMAIN 3467 3511 EGF-LIKE 13.
 FT DOMAIN 3512 3552 LDL-RECEPTOR CLASS A 26.
 FT DOMAIN 3553 3593 LDL-RECEPTOR CLASS A 27.
 FT DOMAIN 3594 3634 LDL-RECEPTOR CLASS A 28.
 FT DOMAIN 3635 3675 LDL-RECEPTOR CLASS A 29.
 FT DOMAIN 3678 3718 LDL-RECEPTOR CLASS A 30.
 FT DOMAIN 3719 3758 LDL-RECEPTOR CLASS A 31.
 FT DOMAIN 3759 3797 LDL-RECEPTOR CLASS A 32.
 FT DOMAIN 3798 3836 LDL-RECEPTOR CLASS A 33.
 FT DOMAIN 3842 3882 LDL-RECEPTOR CLASS A 34.
 FT DOMAIN 3883 3924 LDL-RECEPTOR CLASS A 35.
 FT DOMAIN 3928 3966 LDL-RECEPTOR CLASS A 36.
 FT DOMAIN 3968 4008 EGF-LIKE 14.
 FT DOMAIN 4009 4050 EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 4156 4197 LDL-RECEPTOR CLASS B 35.
 FT DOMAIN 4241 4284 LDL-RECEPTOR CLASS B 36.
 FT DOMAIN 4332 4370 LDL-RECEPTOR CLASS B 37.
 FT DOMAIN 4379 4413 EGF-LIKE 16.
 FT DOMAIN 4414 4454 EGF-LIKE 17.
 FT DOMAIN 4457 4463 SH3-BINDING (POTENTIAL).
 FT DOMAIN 4606 4609 SH2-BINDING (POTENTIAL).
 FT DOMAIN 4619 4625 SH3-BINDING (POTENTIAL).
 FT DOMAIN 4624 4630 SH3-BINDING (POTENTIAL).
 FT SITE 1743 1745 CELL ATTACHMENT SITE (POTENTIAL).
 FT DOMAIN 4522 4527 ENDOCYTOSIS SIGNAL (POTENTIAL).
 FT DOMAIN 4601 4606 ENDOCYTOSIS SIGNAL (POTENTIAL).
 FT DISULFID 28 40 BY SIMILARITY.
 FT DISULFID 35 53 BY SIMILARITY.
 FT DISULFID 47 62 BY SIMILARITY.
 FT DISULFID 67 80 BY SIMILARITY.
 FT DISULFID 74 93 BY SIMILARITY.
 FT DISULFID 87 103 BY SIMILARITY.
 FT DISULFID 108 120 BY SIMILARITY.
 FT DISULFID 115 133 BY SIMILARITY.
 FT DISULFID 127 142 BY SIMILARITY.
 FT DISULFID 147 157 BY SIMILARITY.
 FT DISULFID 152 170 BY SIMILARITY.
 FT DISULFID 164 179 BY SIMILARITY.
 FT DISULFID 183 195 BY SIMILARITY.
 FT DISULFID 190 208 BY SIMILARITY.

Query Match 85.7%; Score 30; DB 1; Length 4660;
 Best Local Similarity 83.3%; Pred. No. 8; Le=02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PROPOG 6
 Db 2635 PROPSG 2640

RESULT 12
 MFAL USTMA STANDARD; PRT; 40 AA.
 AC P31962
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE A1-SPECIFIC PHEROMONE (MATING FACTOR A1).
 GN MFAL.
 OS Ustilago maydis (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 92154661.
 RA Boelker M., Urban M., Kahman R.;
 RT "The a mating type locus of *U. maydis* specifies cell signaling
 components";
 RL Cell 68:441-450(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96186872.
 RA Urban M., Kahman R., Bolker M.;
 RT "The allelic a mating type locus of *Ustilago maydis*: remnants of an
 additional pheromone gene indicate evolution from a multiallelic
 ancestor";
 RL Mol. Gen. Genet. 250:414-420(1996).
 CC -1- FUNCTION: MATING PHEROMONE FOR A1 ALLELE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M84159; AAA34227.1; -;
 DR EMBL: U37795; AAA9765.1; -;
 DR PIR: A42087; A42087.
 KW Prenylation; Lipoprotein; Pheromone.
 FT LIPID
 SQ SEQUENCE 40 AA; 4113 MW; 62D309702F48BC5 CRC64;

Query Match 82.9%; Score 29; DB 1; Length 40;
 Best Local Similarity 83.3%; Pred. No. 9.6;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PRPOG 6
 Db 20 PRPOG 25

RESULT 13
 ID CRA_PLAFA STANDARD; PRT; 162 AA.
 AC P04923:
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CIRCUMSPOROZOITE-PROTEIN RELATED ANTIGEN PRECURSOR (CRA).
 OS Plasmodium falciparum.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85270476.
 RA Coppel R.L., Pavladoro J.M., Grewther P.E., Burkot T.R., Bianco A.E.,
 RA Stahl H.-D., Kemp D.J., Anders R.F., Brown G.V.;
 RT "A blood stage antigen of plasmodium falciparum shares determinants
 RT with the sporozoite coat protein";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:5121-5125(1985).
 CC -1- MISCELLANEOUS: ALTHOUGH THERE ARE NO AUTHENTIC REPEATS IN THIS
 CC (N-A-D-P), THE FIRST OF THESE TETRAPEPTIDES (N-A-N-P) AND
 CC ANTIGEN, THERE ARE A NUMBER OF INTERNAL HOMOLOGIES (N-A-N-P) AND
 CC FOUND IN THE CSP OF P. FALCIPARUM & REACTS WITH ANTIBODIES AGAINST
 CC CRA. IT IS POSSIBLE THAT IMMUNE RESPONSES TO CRA MAY ACT AGAINST
 CC THE CSP ALSO. THE CRA IS FOUND IN MANY PARASITIC STRAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: M11145; AAA29523.1; -;
 DR PIR: A25780; A25780.
 KW Signal; Malaria.
 FT SIGNAL
 FT CHAIN 17 162
 FT CHAIN 17 162
 SQ SEQUENCE 162 AA; 17349 MW; FB60C8250BC0B589 CRC64;

Query Match 82.9%; Score 29; DB 1; Length 162;
 Best Local Similarity 83.3%; Pred. No. 41;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PRPOG 6
 Db 146 PRPOG 151

RESULT 14
 ID EXPL_PLAFA STANDARD; PRT; 162 AA.
 AC P04926; P06718;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MALARIA PROTEIN EXP-1 PRECURSOR (EXPORTED ANTIGEN AG 5.1).
 CC EXP-1.
 GN Plasmodium falciparum.
 OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87218504.
 RA Simmons D., Woolllett G., Bergin-Cartwright M., Kay D., Scalfe J.;
 RT "A malaria protein exported into a new compartment within the host
 RT erythrocyte";
 RL EMBO J. 6:485-491(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85215483.
 RA Hope I.A., McKay M., Hyde J.E., Goman M., Scalfe J.;
 RT "The gene for an exported antigen of the malaria parasite Plasmodium
 RT falciparum cloned and expressed in *Escherichia coli*";
 RL Nucleic Acids Res. 13:369-379(1985).
 CC -1- MISCELLANEOUS: THIS PROTEIN IS ADDRESSED TO A NEW COMPARTMENT
 CC WITHIN THE CYTOPLASM OF THE INFECTED RED CELL. IT ALSO SURROUNDS
 CC THE PARASITE, PROBABLY IN THE PARASITOPHOROUS VACUOLE MEMBRANE.
 CC -1- MISCELLANEOUS: THIS ANTIGEN AND THE CIRCUMSPOROZOITE PROTEIN
 CC APPEAR TO HAVE A COMMON EPITOPE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X05074; CAA28735.1; -;
 DR EMBL: X01745; CAA25881.1; -;
 DR EMBL: A04553; CAA00366.1; -;
 DR PIR: A23052; YAZ051.
 DR PIR: A26769; A26769.
 DE Malaria; Sporozoite; Transmembrane; Signal.
 FT SIGNAL
 FT CHAIN 1 22
 FT CHAIN 23 162
 FT TRANSMEM 80 101
 FT DOMAIN 120 137
 FT VARIANT 136 136
 FT VARIANT 160 160
 FT CONFLICT 10 10
 FT CONFLICT 10 10
 SQ SEQUENCE 162 AA; 17450 MW; AF6B57446E4AA212 CRC64;

Query Match 82.9%; Score 29; DB 1; Length 162;
 Best Local Similarity 83.3%; Pred. No. 41;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PROPOG 6
 Db 146 PROPOG 151

Search completed: November 16, 2000, 15:38:22
 Job time: 108 sec

```

RESULT 15
SSB_SERMA
ID SSB_SERMA STRAND: PRT: 175 AA.
AC P25762.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SINGLE-STRAND BINDING PROTEIN (SSB) (HELIX-DESTABILIZING PROTEIN).
GN SSB.
OS Serratia marcescens
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SR41;
RX MEDLINE: 9352278.
RA de Vries J., Wackernagel W.;
RT "Cloning and sequencing of the Serratia marcescens gene encoding a
RL single-stranded DNA-binding protein (SSB) and its promoter region."
RN Gene 127:39-45(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE: 9501041.
RA de Vries J., Genschel J., Urbanke C., Thole H., Wackernagel W.;
RT "The single-stranded-DNA-binding proteins (SSB) of Proteus mirabilis
RL and Serratia marcescens."
RN Eur. J. Biochem. 224:613-622(1994).
CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR REPLICATION OF THE
CC CHROMOSOMES AND ITS SINGLE-STRAND DNA PHAGES. IT IS ALSO
CC INVOLVED IN DNA RECOMBINATION AND REPAIR.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SSB FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65090; CAA46208.1; -.
DR PIR: JN0632; JN0632.
DR HSSP: P02339; IKAW.
DR INTERPRO: IPR000424; -.
DR PRAM: PF00436; SSB; 1.
DR PROSITE: PS00735; SSB_1; 1.
DR PROSITE: PS00736; SSB_2; 1.
KW DNA-binding; DNA repair; DNA replication.
FT INT_MET 0
FT DNA_BIND 54 60 BY SIMILARITY.
FT SHOUINCE 175 AA; 18678 MW; 05F971ACE9B7239D CRC64;

```

Query Match 82.9%; Score 29; DB 1; Length 175;
 Best Local Similarity 83.3%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PROPOG 6
 Db 135 PROPOG 140

. Thu, Nov 16 15:44:24 2000

us-09-308-237b-1.rsp

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 16, 2000, 15:36:10 : Search time 18.12 Seconds
(without alignments)
30.919 Million cell updates/sec

Title: US-09-308-237b-1
Perfect score: 35
Sequence: 1 PTCPOG 6

Scoring table: BLOSUM62
Gapop: 10.0, Gapext: 0.5

Searched: 297973 seqs, 9337436 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP omc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP protist:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	573	4	016671 homo sapien
2	32	91.4	302	2	092621 chlamydia t
3	32	91.4	601	5	09v746 drosophila
4	32	91.4	792	2	084244 chlamydia t
5	31	88.5	300	13	09v748 gallus gall
6	31	88.5	557	11	062893 rattus norv
7	31	88.5	557	11	063045 rattus norv
8	31	88.5	645	5	09v746 rattus norv
9	31	88.5	780	4	075157 homo sapien
10	31	88.5	3070	12	089906 beet yellow
11	31	88.5	3074	12	008534 sugar beet
12	31	88.6	3090	12	090703 beet yellow
13	30	85.7	213	10	080529 arabidopsis
14	30	85.7	297	13	013015 carassius a
15	30	85.7	298	2	09v740 delinococcus
16	30	85.7	433	5	09v741 drosophila
17	30	85.7	471	13	09v741 salmonella
18	30	85.7	471	13	09v741 xenopus lae
19	30	85.7	603	12	071296 rat cytomeg

20	30	85.7	702	10	09v746	09v746 arabidopsis
21	30	85.7	795	5	09v746	09v746 drosophila
22	30	85.7	858	5	09v746	09v746 bos taurus
23	30	85.7	948	6	077658	077658 bos taurus
24	30	85.7	948	6	077658	077658 bos taurus
25	30	85.7	952	6	077660	077660 bos taurus
26	30	85.7	1160	5	018261	018261 caenorhabd
27	30	85.7	1418	2	09v741	09v741 delinococcus
28	30	85.7	2087	11	09v748	09v748 rattus norv
29	30	85.7	2125	5	096057	096057 halocynthia
30	30	85.7	2158	11	09v743	09v743 rattus norv
31	30	85.7	2167	11	09v748	09v748 rattus norv
32	30	85.7	2295	5	09v742	09v742 caenorhabd
33	30	85.7	2482	5	018263	018263 caenorhabd
34	30	85.7	2635	12	09v743	09v743 rattus norv
35	30	85.7	2635	12	040942	040942 kaposi's sa
36	30	85.7	3375	5	09v743	09v743 rattus norv
37	29	82.9	14	2	09v743	09v743 rattus norv
38	29	82.9	91	5	09v743	09v743 rattus norv
39	29	82.9	99	1	030096	030096 archaeoglob
40	29	82.9	124	3	074492	074492 schizosacch
41	29	82.9	136	2	09v743	09v743 rattus norv
42	29	82.9	141	4	09v743	09v743 rattus norv
43	29	82.9	145	10	09v743	09v743 rattus norv
44	29	82.9	162	5	025840	025840 plasmidium
45	29	82.9	162	5	09v743	09v743 rattus norv

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	573 AA.
ID 016671	016671: 013762:		
AC 01-NOV-1996 (TREMURREL. 01, Created)			
DT 01-NOV-1996 (TREMURREL. 01, Last sequence update)			
DT 01-JUN-2000 (TREMURREL. 14, Last annotation update)			
DE ANTI-MULLERIAN HORMONE TYPE II RECEPTOR PRECURSOR.			
GN MISRL1.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OC [1]			
RP SEQUENCE FROM N.A.			
RA Imbeaud S., Faure E., Lamarre I., Mattei M.G., di Clemente N.,			
RA Tizard R., Carre-Eusebe D., Belleville C., Tragechon L., Tonkin C.,			
RA Nelson J., McAlliffe M., Bidart J.M., Lababidi A., Jossio N.,			
RA Cate R.L., Picard J.Y.,			
RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.			
RL [2]			
RP SEQUENCE FROM N.A.			
RA Imbeaud S., Faure E., Lamarre I., Mattei M.G., di Clemente N.,			
RA Tizard R., Carre-Eusebe D., Belleville C., Tragechon L., Tonkin C.,			
RA Nelson J., McAlliffe M., Bidart J.M., Lababidi A., Jossio N.,			
RA Cate R.L., Picard J.Y.,			
RA "Insensitivity to anti-mullerian hormone due to a mutation in the			
RA human anti-mullerian hormone receptor.",			
RL Nat. Genet. 11:382-388(1995).			
RL [3]			
RP SEQUENCE FROM N.A.			
RA MEDLINE: 96083584.			
RA Visser J.A., McLuskey A., Van Beers T., Weghuis D.O., van Kessel A.G.,			
RA Grootegeed J.A., Thannen A.P.N.,			
RA "Structure and chromosomal localization of the human anti-mullerian			
RA hormone type II receptor gene.",			
RL Biochem. Biophys. Res. Commun. 215:1029-1036(1995).			
RL [4]			
RP SEQUENCE FROM N.A.			
RA TISSUE-TESTIS:			
RC Mastakos P.T., MacLaughlin D.T., Maheswaran S., Teixeira J.M.,			
RC Fuller A.F., Shah P.C., Kehas D.J., Kennally M.K., Dombkowski D.M.,			

Ha T.U., Pfeffer F.I., Donahoe P.K.:
 "Human Ovarian Cancer" Cell Lines and Primary Ascites Cells, Express
 the Human MIS Type II, Bind, and Are Responsive to MIS.";
 CLIN. Cancer Res. 0:0-0(1999).
 DR EMBL: X91156; CAA62593.1; -
 DR EMBL: X91157; CAA62593.1; JOINED.
 DR EMBL: X91158; CAA62593.1; JOINED.
 DR EMBL: X91159; CAA62593.1; JOINED.
 DR EMBL: X91160; CAA62593.1; JOINED.
 DR EMBL: X91161; CAA62593.1; JOINED.
 DR EMBL: X91162; CAA62593.1; JOINED.
 DR EMBL: X91163; CAA62593.1; JOINED.
 DR EMBL: X91164; CAA62593.1; JOINED.
 DR EMBL: X91165; CAA62593.1; JOINED.
 DR EMBL: X91166; CAA62593.1; JOINED.
 DR EMBL: X89013; CAA61418.1; -
 DR EMBL: U29700; AAC50328.1; -
 DR EMBL: AF172932; AAD48497.1; -
 DR INTERPRO: IPR000472; -
 DR INTERPRO: IPR000719; -
 DR PFAM: PF00069; PKINASE; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Signal: Receptor.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 573 ANTI-MULLERIAN HORMONE TYPE II RECEPTOR.
 FT CONFLICT 161 161 L -> V (IN REF. 2).
 SQ SEQUENCE 573 AA; 62749 MW; 1347C10C3942FDBA CRC64;

Query Match
 Best Local Similarity 100.0%; Score 35; DB 4; Length 573;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
 DB 369 PROPOG 374

RESULT 2
 09ZCN2 PRELIMINARY; PRT; 302 AA.
 AC 09ZCN2;
 DT 01-MAY-1999 (TREMBLREL. 10, Created)
 DT 01-MAY-1999 (TREMBLREL. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLREL. 10, Last annotation update)
 DN OUTER MEMBRANE PROTEIN VAEI (PRAGMENT).
 CN VAEI.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-LGV-434;
 RA Hannant J.P., Rockey D.D.:
 RT "Use of a primate model system for identification of Chlamydia
 trachomatis proteins recognized uniquely in the context of
 infection."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF077009; AAC35945.1; -
 FT NON-TER 1 1
 SQ SEQUENCE 302 AA; 33548 MW; 8D297855C17ED6B1 CRC64;

Query Match
 Best Local Similarity 91.4%; Score 32; DB 2; Length 302;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
 DB 202 PROPOG 207

RESULT 3
 09V746 PRELIMINARY; PRT; 601 AA.
 AC 09V746;
 DT 01-MAY-2000 (TREMBLREL. 13, Created)
 DT 01-MAY-2000 (TREMBLREL. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLREL. 14, Last annotation update)
 DE MSPO PROTEIN.
 OS MSPO.
 GN Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RX MEDLINE: 20196006.
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aspayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley S.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fodor C., Gabriellian A.E., Gary N.S., Gelbart M.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaisai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paule J.M.,
 RA Palazzolo M., Pittman K.A., Pan S., Pollard J., Port V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003813; AAP58219.1; -
 DR FLVASE: FB990020269; msps.
 DR INTERPRO: IPR000884; -
 DR INTERPRO: IPR002106; -
 DR PFAM: PF00090; tsp_1; 1.
 DR PROSITE: PS00339; AA:TRNA_LIGASE_II.2; UNKNOWN.1
 SQ SEQUENCE 601 AA; 65434 MW; 529EDCD348AC8BF CRC64;

Query Match
 Best Local Similarity 91.4%; Score 32; DB 5; Length 601;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
 DB 338 PROPOG 343

RESULT 4
ID 084244 PRELIMINARY; PRT; 792 AA.
AC 084244.
DT 01-NOV-1998 (TREMBLER. 08, Created)
DT 01-NOV-1998 (TREMBLER. 08, Last sequence update)
DE 01-MAY-2000 (TREMBLER. 13, Last annotation update)
GN OMP85 ANALOG.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/WR-3/CX;
RA Stephens R.S., Kelman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL: A5001297; AAC67834.1; -
DR INTERPRO: IPR000184; -
DR PFAM: PF01103; Btc_surface_Ag; 1.
SQ SEQUENCE 792 AA; 88751 MW; 90916A8E43E9FDB0 CRC64;

Query Match 91.4%; Score 32; DB 2; Length 792;
Best Local Similarity 83.3%; Pred. No. 98;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
DB 692 PROPOG 697

RESULT 5
ID 09YH8 PRELIMINARY; PRT; 300 AA.
AC 09YH8.
DT 01-MAY-1999 (TREMBLER. 10, Created)
DT 01-MAY-1999 (TREMBLER. 10, Last sequence update)
DE 01-MAY-2000 (TREMBLER. 13, Last annotation update)
GN HOMEODOMAIN PROTEIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 98449697.
RA Tanabe T., Williams C., Jessell T.M.;
RT "Specification of motor neuron identity by the MNR2 homeodomain
protein."
RL Cell 95:67-80(1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AF066860; AAC64924.1; -
DR HSSP: P14653; 1872.
DR INTERPRO: IPR001356; -
DR PFAM: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX; 1.
DR PROSITE: PS00027; HOMEBOX; 1.
DR PROSITE: PS0071; HOMEBOX; 2; 1.
KW Homeobox; DNA-binding; Nuclear protein.
SQ SEQUENCE 300 AA; 32300 MW; 19F710C312EDC3E6 CRC64;

Query Match 88.6%; Score 31; DB 13; Length 300;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6

DB 238 PROPOG 243
RESULT 6
ID 062893 PRELIMINARY; PRT; 557 AA.
AC 062893.
DT 01-NOV-1996 (TREMBLER. 01, Created)
DT 01-NOV-1996 (TREMBLER. 01, Last sequence update)
DE 01-JUN-2000 (TREMBLER. 14, Last annotation update)
GN MIS TYPE II RECEPTOR.
DE RATRII.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 96107132.
RA Teixeira J., He W.W., Shah P.C., Morikawa N., Lee M.M., Catlin E.A.,
RA Hudson P.L., Wing J., MacLaughlin D.T., Donahoe P.K.;
RT "Developmental expression of a candidate muellerian inhibiting
substance type II receptor."
RL Endocrinology 137:160-165(1996).
DR EMBL: U42427; AAC52343.1; -
DR INTERPRO: IPR000472; -
DR INTERPRO: IPR000719; -
DR INTERPRO: IPR001245; -
DR PFAM: PF00069; pkinase; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
SQ SEQUENCE 557 AA; 59748 MW; 8ED99C0C32EBDD5 CRC64;

Query Match 88.6%; Score 31; DB 11; Length 557;
Best Local Similarity 83.3%; Pred. No. 11e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
DB 367 PROPOG 372

RESULT 7
ID 063045 PRELIMINARY; PRT; 557 AA.
AC 063045.
DT 01-NOV-1996 (TREMBLER. 01, Created)
DT 01-NOV-1996 (TREMBLER. 01, Last sequence update)
DE 01-JUN-2000 (TREMBLER. 14, Last annotation update)
GN PUT. ANTI-MULLERIAN HORMONE RECEPTOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS.
RC MEDLINE: 94163972.
RA Baarends W.M., Van Helmond M.J.L., Post M., Van der Schoot P.J.C.M.,
RA Hoogerbrugge J.W., de Winter J.P., Vliegenhart J.T.J., Karels B.,
RA Wilting L.G., Melijers J.H.C., Thammann A.P.N., Grootegeed A.J.;
RT "A novel member of the transmembrane serine/threonine kinase receptor
family is specifically expressed in the gonads and in mesenchymal
cells adjacent to the muellerian duct."
RL Development 120:189-197(1994).
DR EMBL: X71916; CAA50731.1; -
DR INTERPRO: IPR000472; -
DR INTERPRO: IPR000719; -
DR INTERPRO: IPR001245; -
DR PFAM: PF00069; pkinase; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
KW Hormone.

OY 1 PROPOG 6
 1111:1
 Db 1124 PROPG 1129

RESULT 11
 ID 008534 PRELIMINARY: PRT: 3074 AA.

AC 008534; 008534; 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE PUTATIVE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)
 DE (CONTAINS: 66 KDA PROTEIN).
 OS Sugar beet yellow virus (SBYV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
 OC Closterovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-UKRAINIAN (BYV-U);
 RX MEDLINE: 94082464.
 RA Agronomovskiy A.A., Koonin E.V., Boyko V.P., Maiss E., Froetschl R.,
 Lunina N.A., Alatekov J.G.;
 RT "Beet yellows closterovirus: complete genome structure and
 identification of a leader papain-like thiol protease.";
 RL VIROLOGY 198:311-324(1994).
 CC - FUNCTION: PROBABLE POLYMERASE.
 CC - MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY A RIBOSOMAL FRAMESHIFT
 CC THAT OCCURS BETWEEN POSITIONS 2630 AND 2631.
 CC EMBL: X73476; CA51871.1; -.
 DR INTERPRO: IPR000006; -.
 DR INTERPRO: IPR001788; -.
 DR INTERPRO: IPR002388; -.
 DR PFAM: PF00978; RNA_dep_RNAPol2; 1.
 DR PRAM: PF01443; Viral_helicase1; 2.
 DR PRAM: PF01660; Vmethyltransf; 1.
 DR RNA-directed RNA polymerase.
 KW CHAIN 1 3074 348 KDA PROTEIN.
 FT CHAIN 1 588
 FT MUTAGEN 509 509 C->T: NOT PROCESSED INTO ITS SUBUNITS.
 FT MUTAGEN 517 517 C->T: REDUCED PROCESSING INTO SUBUNITS.
 FT MUTAGEN 518 518 C->T: REDUCED PROCESSING INTO SUBUNITS.
 FT MUTAGEN 556 556 H->E: NO EFFECT.
 FT MUTAGEN 569 569 H->E: NOT PROCESSED INTO ITS SUBUNITS.
 FT MUTAGEN 588 588 G->D: NOT PROCESSED INTO ITS SUBUNITS.
 SQ SEQUENCE 3074 AA; 346028 MW; D75B956458BBD0C CRC64;

Query Match 88.6%; Score 31; DB 12; Length 3074;
 Best Local Similarity 83.3%; Pred. No. 6e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 1; Mismatches 0;

OY 1 PROPOG 6
 1111:1
 Db 1124 PROPG 1129

RESULT 12
 ID 090703 PRELIMINARY: PRT: 3090 AA.
 AC 090703; 090703; 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE Beet yellows virus.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
 OC Closterovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BYV-4;
 RX MEDLINE: 20079557.
 RA Peremylov V.V., Hagiwara Y., Dolja V.V.;

RT "HSP70 homolog functions in cell-to-cell movement of a plant virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BYV-4;
 RA Peremylov V.V., Hagiwara Y., Alzhanova D., Dolja V.V.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF190581; AAF14300.1; -.
 DR INTERPRO: IPR000606; -.
 DR INTERPRO: IPR001788; -.
 DR INTERPRO: IPR002588; -.
 DR PFAM: PF00978; RNA_dep_RNAPol2; 1.
 DR PFAM: PF01443; Viral_helicase1; 1.
 DR PFAM: PF01660; Vmethyltransf; 1.
 SQ SEQUENCE 3090 AA; 347990 MW; 09C779A57DE874B9 CRC64;

Query Match 88.6%; Score 31; DB 12; Length 3090;
 Best Local Similarity 83.3%; Pred. No. 6e+02; Indels 0; Gaps 0;
 Matches 5; Conservative 1; Mismatches 0;

OY 1 PROPOG 6
 1111:1
 Db 1124 PROPG 1129

RESULT 13
 ID 080529 PRELIMINARY: PRT: 213 AA.
 AC 080529; 080529; 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE P1439.12 PROTEIN.
 CN P1439.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
 OC Brassicaceae; Arabidopsis.
 OC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
 Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,
 Oj O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,
 Yu G., Becker J., Theologis A., Davis R.W.;
 RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AC003970; AAC33206.1; -.
 DR EMBL: AC003970; AAC33206.1; 4216943C2CD3FBBF CRC64;
 SQ SEQUENCE 213 AA; 22308 MW; 4216943C2CD3FBBF CRC64;

Query Match 85.7%; Score 30; DB 10; Length 213;
 Best Local Similarity 83.3%; Pred. No. 68;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PROPOG 6
 1111:1
 Db 107 PROPG 112

RESULT 14
 ID 013015 PRELIMINARY: PRT: 297 AA.
 AC 013015; 013015; 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE Y BOX PROTEIN 2.
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Cyprinidae; Cyprinidae; Carassius.
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-OVARY:
 RX MEDLINE: 98055714.
 RA Katsu Y., Yamashita M., Nagahama Y.,
 RT "Isolation and characterization of goldfish y box protein, a germ-
 cell-specific RNA-binding protein."
 RL Eur. J. Biochem. 249:854-861(1997).
 CC -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
 DR EMBL: AB003336; BAA19850.1; -.
 DR HSSP: P32081; IMG.
 DR INTERPRO: IPR002059; -.
 DR PFAM: PF00313; CSD; 1.
 DR PRINTS: PR00050; COLD_SHOCK.
 DR PROSITE: PS00352; COLD_SHOCK; 1.
 KW Transcription regulation; DNA-binding.
 SQ SEQUENCE 297 AA: 32775 MM: D006F95C8EB76F33 CRC64;

Query Match 85.7%; Score 30; DB 13; Length 297;
 Best Local Similarity 83.3%; Pred. No. 95;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
 1 1111
 Db 158 PVOPOG 163

RESULT 15
 O9RV70
 ID O9RV70 PRELIMINARY: PRT: 298 AA.
 AC O9RV70:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE URICASE.
 DE DR1160.
 GN OS Delinococcus radiodurans.
 OC Bacteria; Thermus/Delinococcus group; Delinococcales; Delinococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RL:
 RX MEDLINE: 20036896.
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dooson R.J., Hart D.H., Winn M.L., Nelson W.C., Richardson D.L.,
 RA Moritz K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Ulterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith R.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome Sequence of the Radioresistant Bacterium Delinococcus
 RT radiodurans R1."
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001965; AMF10733.1; -.
 DR TIGR: DR1160; -.
 DR INTERPRO: IPR002042; -.
 DR PFAM: PF01014; URICase; 1.
 DR PRINTS: PR00093; URICASE.
 SQ SEQUENCE 298 AA: 34332 MM: 1CCCT945F20F9C56 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 298;
 Best Local Similarity 83.3%; Pred. No. 95;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 PTOPOG 6
 1 1111
 Db 125 PAOPOG 130

Search completed: November 16, 2000, 15:38:10
 Job time: 120 sec

Thu Nov 16 15:44:24 2000

us-09-308-237b-1-1.rag

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model
Run on: November 8, 2000, 14:35:13 ; Search time 15.15 seconds
(without alignments)
13.542 Million cell updates/sec

Title: US-03-308-237B-1
Sequence: 1 PTQPG 6
Scoring table: BLOSUM62
Gapcp 10.0, Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
1	100.0	573	R80633	Clone 3-2 encoded
2	97.4	797	R80633	Chlamydia trachoma
3	91.4	350	Y84616	A chicken motor ne
4	88.6	350	W14532	Human chimeric fuc
5	88.6	350	W14532	A glycosyltransfer
6	88.6	374	W13642	Human alpha(1,3)-f
7	85.7	20	Y20992	Human glial fibril
8	82.9	13	R28833	W11243 cyclic pep
9	82.9	41	W42445	Ustilago maydis MF
10	82.9	74	P50761	Antigenic protein
11	82.9	125	Y21354	Human Hsp-1 mutan
12	82.9	127	Y73845	Human prostate tum

13	82.9	135	Y44649	Streptomyces grise
14	82.9	162	P51201	Antigenic protein
15	82.9	226	W13338	Group B streptococ
16	82.9	326	W13695	Haemophilus influe
17	82.9	369	W70224	Human RNA-associat
18	82.9	459	W56738	Orp1omyces cellu
19	82.9	460	Y94991	Human secreted pro
20	82.9	485	W65462	Neocallimastix pat
21	82.9	485	Y27283	N. patriciarum xyl
22	82.9	567	W78489	Human p300 ligand
23	82.9	747	W60487	Human TRIDENT tran
24	82.9	755	W81365	Human prothromone c
25	82.9	763	W60591	Human hepatocyte n
26	82.9	1078	R28963	Notch hnk3 full le
27	82.9	1093	R41001	Human myotonic dys
28	82.9	1135	Y94923	Human secreted pro
29	82.9	1661	W76041	Hydra head activat
30	82.9	2414	R84882	Transcription fact
31	82.9	2414	W40057	Cellular transcrip
32	80.0	63	Y30717	Amino acid sequenc
33	80.0	120	Y48601	Human breast tumou
34	80.0	144	Y29974	Human TSC-22 homol
35	80.0	344	R93797	Protein Lp40, a va
36	80.0	400	P70654	Sequence encoded b
37	80.0	405	Y70440	Human Notch signal
38	80.0	531	Y70439	Human Notch signal
39	80.0	547	W40112	Human alpha-6(IV)
40	80.0	561	Y34560	Porphyromonas ging
41	80.0	601	Y34423	Porphyromonas ging
42	80.0	620	Y70438	Human Notch signal
43	80.0	622	Y70437	Human Notch signal
44	80.0	694	R14666	Truncated HSVab po
45	80.0	795	W72062	HSV-2 strain SB5 C

ALIGNMENTS

RESULT 1	
ID	R80633 standard; Protein: 573 AA.
AC	R80633:
DT	27-FEB-1996 (first entry)
DE	Clone 3-2 encoded protein.
XX	Anti-Mullerian hormone receptor; anti-Mullerian hormone; AMH;
KW	receptor; antibody; therapy; HI; 2B10; tumour;
KW	transforming growth factor-beta; TGF-beta; bone morphogenic protein;
KW	activin; Mullerian inhibiting substance; MIS.
OS	Homo sapiens.
XX	
XX	Location/Qualifiers
FT	1..17
FT	/note="signal peptide"
FT	18..144
FT	/note="mature protein"
FT	18..144
FT	/note="extracellular domain"
FT	66..68
FT	/note="potential N-linked glycosylation site"
FT	119..121
FT	/note="potential N-linked glycosylation site"
FT	144..170
FT	/note="transmembrane region"
FT	171..573
FT	/note="cytoplasmic domain"
FT	18..573
FT	/note="claimed"

PN W09516709-A2.
 XX
 PD 22-JUN-1995.
 XX
 PF 13-DEC-1994: 94WO-US14643.
 XX
 PR 23-DEC-1993: 93US-0173512.
 XX
 PR 13-DEC-1993: 93US-0166333.
 XX
 PA (BIOI) BIOGEN INC.
 XX (INRM) INSERM INST NAT SANTE & RECH MED.
 PI Gate RL, Josso N;
 XX
 DR WPI: 1995-231521/30.
 DR N-PSDB: Q98988.
 XX
 PT New DNA encoding anti-Mullerian hormone receptor, related
 PT polypeptide(s) and antibodies - useful in immunosays screening for
 XX ligands, treatment of cancer cells expressing the receptor etc.
 PS
 XX Claim 4: Page 54-56; 77pp: English.
 CC
 CC This sequence represents the protein encoded by the cDNA clone 3-2. This
 CC sequence was found to be the human AMH-receptor. AMH is a glycoprotein
 CC and is part of the transforming growth factor-beta superfamily. AMH is
 CC structurally and functionally related to TGF-beta bone morphogenic
 CC protein, and activin. AMH receptors are present in a limited number of
 CC tissues and can therefore be used to design antibody-toxin complexes to
 CC target tumour cells in these tissues. The AMH-receptor sequences can be
 CC used to screen and purify compounds that bind to them, thereby
 CC identifying AMH ligands. The antibody designed from the receptor
 CC sequences can be used in immunoassays to detect the levels of
 CC AMH-receptor. The antibody can also be linked to a toxin in order to
 CC kill cells that express the receptor, e.g. cancer cells. The advantage
 CC of using antibody conjugates based on AMH-receptors are that they can be
 CC used more aggressively than those based on antibodies that recognise
 CC receptors more generally distributed.
 XX
 SQ Sequence 573 AA;

Query Match
 Best Local Similarity 100.0%; Score 35; DB 16; Length 573;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PROPOG 6
 DB 369 pcpqpg 374

RESULT 2
 ID Y36955 standard; Protein; 797 AA.
 XX
 AC Y36955;
 XX
 DT 07-OCT-1999 (first entry)
 XX
 DE Chlamydia trachomatis surface exposed protein.
 KW Vaccines; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; peritrichitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW Bartholinitis; pneumonia; venereal lymphogranulomatosis.
 OS Chlamydia trachomatis.
 XX
 PN W09928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 27-NOV-1998: 98WO-IB01939.

XX
 PR 04-NOV-1998: 98US-0107077.
 PR 28-NOV-1997: 97ER-0015041.
 PR 17-DEC-1997: 97ER-0016034.
 XX
 PA (GERT) GENSET.
 XX
 PI Griffiths R;
 XX
 DR WPI: 1999-371125/31.
 XX
 PT Genome sequence of Chlamydia trachomatis
 XX
 PS Disclosure; Page 802; 1755pp: English.
 CC Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
 CC Chlamydia trachomatis (see 201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC Bartholinitis, Bartholinitis; pneumonia; venereal lymphogranulomatosis.
 CC may be of use in treating these diseases.
 XX
 SQ Sequence 797 AA;

Query Match
 Best Local Similarity 91.4%; Score 32; DB 20; Length 797;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PROPOG 6
 DB 697 ptepg 702

RESULT 3
 ID Y84616 standard; Protein; 300 AA.
 XX
 AC Y84616;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE A chicken motor neuron restricted pattern 2 protein.
 KW Motor neuron restricted pattern 2 protein.
 KW Pax6+ motor neuron progenitor; somatic motor neuron;
 KW transcription factor; Isl2; Lim 3; homeobox protein 9; HB9;
 KW chronic neurodegenerative disease; spinal muscular atrophy;
 KW myotrophic lateral sclerosis; neuromuscular disease; sacral agenesis.
 OS Gallus sp.
 XX
 PN W020001884-A1.
 XX
 PD 06-APR-2000.
 XX
 PF 29-SEP-1999: 99WO-US22517.
 XX
 PR 29-SEP-1998: 98US-0162524.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Jesseil TM, Tanabe Y, William C;
 XX
 DR WPI: 2000-293133/25.
 DR N-PSDB: Z99988.
 XX
 PT New isolated nucleic acid molecules encoding motor neuron restricted

PT pattern protein and a homeobox protein are used in treating chronic
 PT neurodegenerative diseases and acute nervous system injuries -
 PS Claim 31; Fig 10; 209pp; English.
 XX
 CC The present sequence represents a motor neuron restricted pattern (MNR2)
 CC protein from a chicken embryo. MNR2 is expressed selectively by Pax6+
 CC motor neuron progenitors and persists in post-mitotic somatic motor
 CC neurons. Expression of the MNR2 protein induces expression of
 CC transcription factors Isl2, Isl3 and homeobox protein 9 (Hb9).
 CC MNR2 and Hb9 are used to induce differentiation of somatic motor
 CC neurons in a mammal which is to treat an abnormality associated with a
 CC lack of one or more normally functioning motor neurons, a chronic
 CC neurodegenerative disease such as spinal muscular atrophy or myotrophic
 CC lateral sclerosis, an acute nervous system injury localized to a
 CC specific central axon and neuromuscular disease. Hb9 is also used to
 CC treat an embryo afflicted with sacral agenesis.
 XX
 SQ Sequence 300 AA;
 OY 1 PROPOG 6
 DB 238 pspgpg 243
 Query Match 88.6%; Score 31; DB 21; Length 300;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 RESULT 4
 ID W14532 standard; Protein: 350 AA.
 XX
 AC W14532;
 XX
 DT 19-JUN-1997 (first entry)
 XX
 DE Human chimeric fucosyltransferase Fuc-TC21.
 XX
 KW Alpha(1,3/1,4)-fucosyltransferase; Lewis enzyme; Fuc-TIII;
 KW alpha(1,3)-fucosyltransferase; Fuc-TV; glycosylation;
 KW oligosaccharide; Fuc-TC21.
 XX
 OS Synthetic.
 XX
 FH key location/Qualifiers
 FT Region 1..99 /label= Fuc-TIII
 FT /note= "amino acids 1-99 of Fuc-TIII"
 FT Region 100..150 /label= Fuc-TV
 FT /note= "amino acids 116-166 of Fuc-TV"
 FT Region 151..351 /label= Fuc-TIII
 FT /note= "amino acids 161-361 of Fuc-TIII"
 FT
 PN WO9709421-A1.
 XX
 PD 13-MAR-1997.
 XX
 PF 06-SEP-1996; 96WO-US13816.
 XX
 PR 08-SEP-1995; 95US-0525058.
 XX
 RA (UNMI) UNIV MICHIGAN.
 XX
 PL Legault DJ, Lowe JB;
 XX
 DR WPI; 1997-192897/17.
 XX
 PT New recombinant fucosyltransferase proteins - useful for modifying
 PT cell surface oligosaccharide structures

XX
 PS Claim 1; Refer to Page 298; 329pp; English.
 XX
 CC Chimeric fucosyltransferases Fuc-TC1 to Fuc-TC21 (W14512-32)
 CC comprise portions of the human GDP-Fuc:(beta-D-Gal(1,4/1,3))-D-
 CC GlcNAc(6/6) alpha(1,3/1,4)-fucosyltransferase (Fuc-TIII, Lewis
 CC enzyme) (W13638) and portions of the GDP-Fuc:(beta-D-Gal(1,4)-D-
 CC GlcNAc alpha(1,3))fucosyltransferase (Fuc-TVI) (W13643) or Fuc-TV
 CC (W13642). They are obtained by cassette mutagenesis of Fuc-TIII and
 CC Fuc-TVI or Fuc-TV nucleic acids (T61675, T61679-80) and expression
 CC in transformed host cells. The chimeric fucosyltransferases can
 CC be used to modify cell surface oligosaccharide structures.
 XX
 SQ Sequence 350 AA;
 OY 1 PROPOG 6
 DB 120 ptpgpg 125
 Query Match 88.6%; Score 31; DB 18; Length 350;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 RESULT 5
 ID R45939 standard; Protein: 374 AA.
 XX
 AC R45939;
 XX
 DT 26-JUL-1994 (first entry)
 XX
 DE A glycosyltransferase.
 XX
 KW Glycosyltransferase; fucosyltransferase; GDP-Fuc; in vitro; cell;
 KW surface; oligosaccharide.
 XX
 OS Homo sapiens.
 XX
 PN WO9402616-A.
 XX
 PD 03-FEB-1994.
 XX
 PF 20-JUL-1993; 93WO-US06703.
 XX
 PR 20-JUL-1992; 92US-0914281.
 XX
 RA (UNMI) UNIV MICHIGAN.
 XX
 PI Lowe JB;
 XX
 DR WPI; 1994-048874/06.
 DR N-PSDB; Q56911.
 XX
 PT DNA fragment encoding a glycosyltransferase - can be used for in
 PT vitro reactions to modify cell surface oligosaccharide(s) e.g.
 PT blood gp. determinants, to protect against transplant rejection
 XX
 PS Disclosure: Fig 6; 249pp; English.
 XX
 CC The sequence is that of a human glycosyl transferase. The enzyme
 CC may be non glycosylated. This prevents premature loss of enzyme
 CC activity. It can also be used in in vitro reactions to modify cell
 CC surface oligosaccharide moles. e.g. blood group determinants.
 CC See also R45933-8.
 XX
 SQ Sequence 374 AA;
 OY
 DB
 Query Match 88.6%; Score 31; DB 15; Length 374;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 11 1111
14 p1spg 19

RESULT 8
R28833 standard; peptide; 13 AA.
XX
AC R28833;
XX
DT 26-APR-1993 (first entry)
XX
DE WF11243 cyclic peptide.
XX
KM Antifungal; antiprotazoal; antibiotic; cyclic.
XX
OS FERM BP-3373.
XX
FH Key
FT Modified-site 1 Location/Qualifiers
FT /note= "N-terminal acylated with X(13), i.e.
FT 3-hydroxy-hexadecanoic acid"
FT 5
FT /label= 4Hyp
FT 8
FT /label= 3Hyp
FT 9
FT /note= "3-hydroxy-Gln"
FT 12
FT /label= Orn
FT 13
FT /label= OTHER
FT /note= "3-hydroxy-hexadecanoic acid, as amino acid
FT analogue; in which the OH group forms an ester bond
FT with Orn(12) and the COOH group forms a peptide
FT bond with Thr(1)"
XX
XX WO9219648-A.
XX
XX 12-NOV-1992.
XX
XX 08-MAY-1992; 92MO-JP00586.
XX
XX 09-MAY-1991; 91JP-0132234.
XX
XX 21-OCT-1991; 91JP-0299552.
XX
XX (FUJI) FUJISAWA PHARM CO LTD.
XX
XX Fujie A, Hashimoto S, Nakanishi T, Okuhara M, Takase S;
XX Yamashita M;
XX
XX WPI; 1992-398800/48.
XX
XX
XX New WF11243 polypeptide - has antifungal and antiprotazoal
XX activity
XX
XX
XX Disclosure; Page 18; 34pp; Japanese.
XX
XX WF11243 is a cyclic polypeptide having antifungal activity (e.g.
XX against Candida albicans, C. glabrata, C. krusei, C. tropicalis,
XX Cryptococcus neoformans, Aspergillus fumigatus and A. niger) and
XX antiprotazoal activity (e.g. against Pneumocystis carinii). It is
XX prepared by culturing FERM BP-3373 and subjecting the culture medium
XX to HPLC.
XX
XX The compound is claimed according to its physicochemical properties:
XX
XX the present sequence is a postulated structure given in the
XX disclosure.
XX
XX
XX Sequence 13 AA;
SQ

Query Match 82.9%; Score 29; DB 13; Length 13;
Best Local Similarity 83.3%; Pred. No. 16;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 PROPG 6
11 111
Db 5 p1spg 10

RESULT 9
W42445 standard; peptide; 41 AA.
XX
AC W42445;
XX
DT 10-JUN-1998 (first entry)
XX
DE Ustilago maydis MFalpal mating pheromone used in a transgenic plant.
XX
DE Ustilago maydis.
XX
KM Fungal pheromone; mating; transgenic plant; fungal infection;
XX
XX Ustilago maydis.
XX
OS Ustilago maydis.
XX
FH Key Location/Qualifiers
FT Cleavage-site 27..28
FT /note= "predicted peptide cleavage site involved
FT in pheromone maturation"
XX
XX WO9748719-A1.
XX
XX 24-DEC-1997.
XX
XX 17-JUN-1997; 97MO-US10364.
XX
XX 17-JUN-1996; 96US-0019598.
XX
XX (TEXA) UNIV TEXAS A & M SYSTEM.
XX
XX Beckerman JL, Ebohle DJ, Kazmierczak P, McCabe P;
XX Van Alfen NJ, Zhang L;
XX
XX WPI; 1998-063072/06.
XX
XX Transgenic plant, whose genome contains a fungal pheromone gene -
XX has increased resistance to fungal infections
XX
XX
XX Example 5; Fig 1; 14pp; English.
XX
XX This is a MFalpal mating pheromone from Ustilago maydis used in the
XX preparation of a transgenic plant with increased resistance to fungal
XX infections. The transgenic plant's genome is augmented by the
XX introduction of a fungal pheromone gene, which confers resistance to
XX fungal infection. The transgenic plant is prepared by a process which
XX comprises obtaining a polynucleotide composition comprising 1 or more
XX fungal pheromone genes, contacting recipient plant cells with the
XX polynucleotide composition, regenerating plants from recipient cells
XX which have received the polynucleotide composition and identifying a
XX fertile, transgenic plant whose genome has been augmented relative to
XX that of the corresponding non-transgenic recipient cells through the
XX stable introduction of the polynucleotide composition. The invention
XX provides for the identification of DNA segments encoding the fungal
XX pheromones, and compositions comprising pheromone polypeptides, which
XX inhibit fungal growth and development and prevent or reduce fungal
XX infection and/or disease in plant host cells. Also provided are methods
XX and processes utilising fungal pheromone and pheromone derivatives in the
XX prevention and treatment of fungal colonisation of a plant.
XX
XX
XX Sequence 41 AA;
SQ

Query Match 82.9%; Score 29; DB 19; Length 41;
Best Local Similarity 83.3%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 DB 20 PTPP9 25

RESULT 10

ID P50761 standard; Protein: 74 AA.

AC P50761;

DT 08-NOV-1991 (first entry)

DE Antigenic protein encoded by clone lambdaag5.1(8).

KW Malaria; vaccine; antigen; epitope.

OS Plasmodium falciparum K1 isolate (Thailand).

XX Key Location/Qualifiers

FT Peptide 32..49 /label= epitope 5.1-1

XX /note= "repeat: see comments"

PN EP153188-A.

PD 28-AUG-1985.

PF 19-FEB-1985; 85EP-0301110.

XX 21-DEC-1984; 84GB-0032337.

PR 21-FEB-1984; 84GB-0004493.

PR 13-NOV-1984; 84GB-0028643.

XX (MATER) NATIONAL RES DEV CORP.

PI Hope IA, Scaife JC, Strambachova-McBride J;

XX WPI: 1985-211745/35.

DR N-PSDB; N50553.

XX Antigenic material for vaccination against malaria - comprises

PT amino acid residues sequence of natural antigen and

PT intra-erythrocytic forms of plasmodium parasites.

XX Disclosure: Fig 5a: 70pp; English.

XX The sequence was deduced from a DNA insert from clone

CC lambdaag5.1(8) which was prep. from RNA from the erythrocytic

CC stages of the parasite. The antigen is present as a sporozoite

CC surface epitope. The 5.1-1 epitope comprises a sequence homologous

CC with the tandemly repeated tetramer present in the circumsporozoite

CC protein (MAMP). The 5.1-1 epitope appears to be made up of two

CC sequences of seven and eight residues respectively linked by three

CC residues, Ser-Glu-Ser, which do not contribute to antigenicity and

CC may be omitted from synthetic antigens. The DNA can be used to

CC express antigenic peptides useful for the prep. of vaccines for

CC protection against malaria. See also P51201.

XX Sequence 74 AA;

Query Match 82.9%; Score 29; DB 6; Length 74;

Best Local Similarity 83.3%; Pred. No. 91;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPPOG 6

DB 58 PTPP9 63

RESULT 11

Y21354

ID Y21354 standard; Protein: 125 AA.

AC Y21354;

DT 22-JUL-1999 (first entry)

DE Human HUPF-1 mutant protein fragment 6.

KW Human; beta-amyloid precursor protein; beta-ApP; diagnosis; cancer;

KW frameshift mutation; age-related disease; neurodegenerative disorder;

KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;

KW Huntington's disease; multiple sclerosis; alcoholic liver disease;

KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;

KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;

KW neurofilament-F; presenilin 1; presenilin II; cellular tumour antigen;

KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;

KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMG-C; NSF-A;

KW high mobility group protein-C; neuroendocrine specific protein A.

XX Synthetic.

OS Homo sapiens.

PN WC0945322-A2.

PD 15-OCT-1998.

PF 02-APR-1998; 98WC-1B00705.

XX 10-APR-1997; 97US-0043163.

XX (UYVY-) RIJMSUNIV UTRECHT.

PA (ROYA-) ROYAL NETHERLANDS ACADEMY OF SCI.

PA (UYRO-) UNIV ROTTERDAM ERASMUS.

XX Burbach JPH, Grosveld FG, Van Leeuwen FW;

XX WPI: 1998-609901/51.

DR N-PSDB; X75768.

XX Diagnosing disease by detecting frameshift mutations in RNA or

PT corresponding protein mutations - used to diagnose cancer and

PT neurological diseases, particularly Alzheimer's disease, and also

PT for treatment and prevention with specific ribozymes or wild-type

XX RNA

XX Disclosure: Figure 17; 258pp; English.

XX This invention describes a novel method for the diagnosis of a disease

CC caused by, or associated with, an RNA molecule that has a frameshift

CC mutation. The method is used to diagnose age-related diseases, especially

CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's

CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,

CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II

CC and many others listed) or susceptibility to these disorders. The method

CC allows a definitive diagnosis of Alzheimer's disease in living patients,

CC at an early stage. It is based on the observation that disease may be

CC caused by mutations in RNA rather than DNA. The invention describes the

CC use of neuronal system RNA molecules, specifically proteins including

CC beta-amyloid precursor protein (beta-ApP), the microtubule associated

CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule

CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,

CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic

CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma

CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group

CC protein-C (HMG-C) and neuroendocrine specific protein A.

XX Sequence 125 AA;

Query Match 82.9%; Score 29; DB 19; Length 125;

Best Local Similarity 83.3%; Pred. No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PROPOG 6
 1 1111
 Db 14 pepqpg 19

RESULT 12
 ID V73845
 V73845 standard; Protein: 127 AA.
 AC V73845;
 DE 14-MAR-2000 (first entry)

DE Human prostate tumor EST fragment derived protein #32.
 KW Pancreas; tumor; EST; expressed sequence tag; human; cytosolic;
 treatment.

OS Homo sapiens.
 PN DEL9820190-A1.
 PD 04-NOV-1999.

PE 28-APR-1998; 98DE-1020190.
 PR 28-APR-1998; 98DE-1020190.
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Spacht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 WPI: 1999-621386/54.
 DR N-PSDB; 252868.

PT New human nuclei: acid sequences from pancreatic tumors, and related
 proteins -

PS Claim 23; Page 321; 502pp; German.

CC This invention describes novel polypeptides and their encoding nucleic
 acids derived from human pancreatic tumor tissue which have cytostatic
 activity. The sequences are also useful in producing pharmaceutical
 compositions for treatment of pancreatic tumors; V73814-V74252 represent
 CC protein fragments encoded by the human pancreatic tumor cDNA library
 CC derived expressed sequence tag (EST) sequences represented in
 CC 252858-253014.

SQ Sequence 127 AA;

Query Match 82.9%; Score 29; DB 20; Length 127;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PROPOG 6
 1 1111
 Db 18 pepqpg 23

RESULT 13
 ID V44649
 V44649 standard; Protein: 135 AA.
 AC V44649;
 DE 18-APR-2000 (first entry)

DE Streptomyces griseus Ssga protein.
 KW Ssga; liquid culture; filamentous bacteria; secondary metabolite;
 mycelium; antibiotic; antitumor agent; immunosuppressive agent;
 hypocholesterolaemic agent; enzyme inhibitor; antimigraine agent;

KW herbicide; antiparasitic agent; ruminant growth promoter;
 KW bioinsecticide; receptor agonist; antagonist; biomass.
 OS Streptomyces griseus.

PN WO200000613-A1.

PD 06-JAN-2000.
 PE 25-JUN-1999; 99MO-NL00395.

PR 26-JUN-1998; 98EP-0202148.

PA (UYLE-) RIJKSUNIV LEIDEN.
 PA (NEME-) NEDERLANDSE ORG WETENSCHAPPELIJK ONDERZO.

PI Van Wezel GP, Kraal B, Luiten RGM;

DR WPI: 2000-147269/13.
 DR N-PSDB; 249727, 249728.

PT Reducing branching and enhancing fragmentation in filamentous
 microorganisms used to improve their liquid culturing properties -
 Disclosure; Fig 5; 60pp; English.

CC The present sequence is S. griseus Ssga protein. Ssga reduces branching
 and fragment separation and enhances fragmentation of mycelium in liquid
 CC culture resulting in lower viscosity of culture broths. Filamentous
 CC bacteria can be transformed with ssga gene-containing plasmid to enhance
 CC the production of secondary metabolites such as, antibiotics, antitumor
 CC agents, immunosuppressive agents, hypocholesterolaemic agents, enzyme
 CC inhibitors, antimigraine agents, herbicides, antiparasitic agents,
 CC ruminant growth promoters, bioinsecticides, receptor agonists and
 CC antagonists and biomass.

SQ Sequence 135 AA;

Query Match 82.9%; Score 29; DB 21; Length 135;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 PROPOG 6
 1 1111
 Db 65 ptepg 70

RESULT 14
 ID P51201
 P51201 standard; Protein: 162 AA.
 AC P51201;

DE 08-NOV-1991 (first entry)

DE Antigenic protein encoded by clone lambda daag5.1(9).
 KW Malaria; vaccine; antigen; epitope.
 OS Plasmodium falciparum K1 isolate (Thailand).

OS Plasmodium falciparum K1 isolate (Thailand).

Key Location/Qualifiers

FT Peptide 1..22

FT Protein 23..162

FT Peptide 120..137

FT Peptide 120..137

FT Region 80..101

FT Region 80..101

FT Region 80..101

FT Region 80..101

FT Region 80..101

FT Region 80..101

FT Region 80..101

FT Region 80..101

```

XX  EP153188-A.
PN
XX  28-AUG-1985.
PD
XX  19-FEB-1985: 85EP-0301110.
PF
XX  21-DEC-1984: 84GB-0032337.
PR  21-FEB-1984: 84GB-0004493.
PR  13-NOV-1984: 84GB-0028643.
XX
PA  (NATR ) NATIONAL RES DEV CORP.
PI
XX  Hopo 1A, Scalfe JG, Strambachova-McBride J;
DR  WPI: 1985-211745/35.
DR  N-PSDB: N50851.
XX
XX  Antigenic material for vaccination against malaria - comprises
PT  amino acid residues sequence of natural antigen and
PI  intra-erythrocytic forms of plasmodium parasites.
XX
PS  Claim 13: Flg 5a; 70pp: English.
XX
CC  The sequence was deduced from a DNA insert from clone lambdaAg5.1(9)
CC  which was prep'd. from RNA from the erythrocytic stages of the
CC  parasite. The antigen is present as a sporozoite surface epitope.
CC  The 5.1-1 epitope comprises a sequence homologous with the tandemly
CC  repeated tetramer present in the circumsporozoite protein (NANP).
CC  The 5.1-1 epitope appears to be made up of two sequences of seven
CC  and eight residues respectively linked by three residues, Ser-Glu-
CC  Ser, which do not contribute to antigenicity and may be omitted
CC  from synthetic antigens. The DNA can be used to express antigenic
CC  peptides useful for the prepn. of vaccines for protection against
CC  malaria. See also P50761.
XX
SQ  Sequence 162 AA:

```

```

Query Match      82.9%; Score 29; DB 6; Length 162;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY  1 PROPG 6
DB  146 pcpqg 151

```

```

RESULT 15
Y91338
ID  Y91338 standard; Protein; 262 AA.
XX
XX  Y91338;
AC
XX
XX  30-MAY-2000 (first entry)
DT
XX
XX  Group B Streptococcus protein sequence SEQ ID NO:71.
DE
XX
XX  Group B Streptococcus; Streptococcus agalactiae; protein antigen;
KW  vaccine; screening; immunogen; detection; diagnosis; infection;
XX  antibody; antibody; antibacterial.
XX
XX  Streptococcus agalactiae.
OS
XX
XX  MO200006736-A2.
PN
XX  10-FEB-2000.
PD
XX
XX  27-JUL-1999; 99MO-GB02444.
PF
XX  27-JUL-1998; 98GB-0016335.
PR  19-MAR-1999; 99US-0125163.
XX

```

```

PA  (MICR-) MICROBIAL TECHNIQS LTD.
XX
XX  Le Page RWF, Wells JM, Hanniffy SB;
PI
XX  WPI: 2000-195299/17.
XX
XX  New Group B Streptococcus protein, useful as vaccine, for diagnosis of
PT  Streptococcal infections and for screening of antibodies or antibodies
XX
XX  Claim 1: Fig 1; 123pp: English.
XX
XX  A05803 to A05872 encode proteins, polypeptides and peptides (given in
CC  as Streptococcus agalactiae. The GBS polynucleotides and polypeptides
CC  have antibacterial activity. Immunogenic compositions comprising GBS
CC  polynucleotides or polypeptides can be used as vaccines and for the
CC  treatment or prophylaxis of GBS infection. The polynucleotides and
CC  polypeptides can also be used in the detection of GBS and for screening
CC  DNA encoding bacterial cell envelope associated or secreted antigens in
CC  gram positive bacteria. A05873 to A05941 represent primers used in the
XX  exemplification of the present invention.
XX
SQ  Sequence 262 AA:

```

```

Query Match      82.9%; Score 29; DB 21; Length 262;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY  1 PROPO 5
DB  26 pcpqg 30

```

Search completed: November 8, 2000, 14:38:16
Job time: 183 sec

Thu, Nov 16 15:44:24 2000

us-09-308-237b-1_1.rag

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2000, 14:35:13 : Search time 12.24 Seconds
(without alignments)
8.216 Million cell updates/sec

Title: US-09-308-237b-1
Perfect score: 35
Sequence: 1 PIR/PGC 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents_AA*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	88.6	374	1 US-07-914-281-11	Sequence 11, Appl
2	31	88.5	374	1 US-08-393-246-11	Sequence 11, Appl
3	31	88.5	374	1 US-08-525-058A-11	Sequence 11, Appl
4	31	88.5	374	1 US-08-696-731-11	Sequence 11, Appl
5	29	82.3	162	2 US-08-319-704-6	Sequence 6, Appl
6	29	82.3	327	2 US-08-651-818A-3	Sequence 6, Appl
7	29	82.3	459	3 US-09-118-319-6	Sequence 3, Appl
8	29	82.3	485	2 US-08-749-391-2	Sequence 2, Appl
9	29	82.3	516	2 US-08-676-166A-7	Sequence 2, Appl
10	29	82.3	755	3 US-09-071-101-2	Sequence 2, Appl
11	29	82.3	755	3 US-09-369-618-2	Sequence 2, Appl
12	29	82.3	753	2 US-08-742-753-4	Sequence 4, Appl
13	29	82.3	1064	1 US-08-537-210A-3	Sequence 3, Appl
14	29	82.3	1068	1 US-08-537-210A-2	Sequence 3, Appl
15	29	82.3	1078	1 US-08-264-534-32	Sequence 11, Appl
16	29	82.3	1078	1 US-08-083-590A-11	Sequence 11, Appl
17	29	82.3	1078	1 US-08-465-500-32	Sequence 32, Appl
18	29	82.3	1078	1 US-08-346-128-32	Sequence 32, Appl
19	29	82.3	1078	3 US-08-532-384-11	Sequence 11, Appl
20	29	82.3	1078	3 US-08-893-828-32	Sequence 32, Appl
21	29	82.3	2414	1 US-08-227-536-2	Sequence 2, Appl
22	29	82.3	2414	4 PCR-US95-04682-2	Sequence 2, Appl
23	29	82.3	2523	1 US-08-185-432-18	Sequence 18, Appl
24	29	82.3	2556	1 US-08-083-580A-20	Sequence 20, Appl
25	29	82.3	2556	1 US-08-532-384-20	Sequence 20, Appl
26	29	80.0	143	2 US-08-889-337-3	Sequence 3, Appl
27	29	80.0	143	2 US-08-889-337-3	Sequence 3, Appl
28	29	80.0	143	3 US-08-851-190-6	Sequence 6, Appl

29	28	80.0	144	2 US-08-889-337-1	Sequence 1, Appl
30	28	80.0	144	3 US-08-851-190-7	Sequence 7, Appl
31	28	80.0	547	1 US-08-494-168-7	Sequence 7, Appl
32	28	80.0	582	3 US-08-906-865-3	Sequence 3, Appl
33	28	80.0	903	1 US-08-220-151-8	Sequence 8, Appl
34	28	80.0	903	1 US-08-413-118-8	Sequence 8, Appl
35	28	80.0	903	3 US-08-804-439A-22	Sequence 22, Appl
36	28	80.0	903	3 US-08-473-446-8	Sequence 8, Appl
37	28	80.0	903	3 US-08-720-229-22	Sequence 2, Appl
38	28	80.0	904	4 PCT-US96-05316-1	Sequence 1, Appl
39	28	80.0	904	4 PCT-US96-05316-2	Sequence 1, Appl
40	28	80.0	904	5 5244792-3	Patent No. 5244792
41	28	80.0	904	5 5244792-4	Patent No. 5244792
42	28	80.0	1243	2 US-08-557-139-2	Sequence 2, Appl
43	28	80.0	1694	1 US-08-494-168-2	Sequence 2, Appl
44	26	74.3	105	1 US-08-417-460-6	Sequence 6, Appl
45	26	74.3	301	4 PCT-US95-13975-72	Sequence 72, Appl

ALIGNMENTS

```

RESULT 1
US-07-914-281-11
; Sequence 11, Application US/07914281
; Patent No. 5324663
;
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOPOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCLELLAND, WAIR & NEUSTADT,
; ADDRESS: P. C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/914,281
; FILING DATE: 19920720
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavelle, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
;
; US-07-914-281-11

```

Query Match 88.6%; Score 31; DB 1; Length 374;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
1 PROPGC 6
|||||

DB 137 PTRPOG 142

RESULT 2
US-08-393-246-11
Sequence 11, Application US/08393246

Patent No. 595900

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/393,246

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/220,433

FILING DATE: 30-MAR-1994

APPLICATION NUMBER: US 07/914,281

FILING DATE: 20-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 374 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-393-246-11

Query Match 88.6%; Score 31; DB 1; Length 374;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTRPOG 6

DB 137 PTRPOG 142

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/525,058A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 374 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-525-058A-11

Query Match 88.6%; Score 31; DB 1; Length 374;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTRPOG 6

DB 137 PTRPOG 142

RESULT 4

US-08-696-731-11

Sequence 11, Application US/08696731

Patent No. 5955347

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/696,731

FILING DATE: 14-AUG-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/393,246

FILING DATE:

APPLICATION NUMBER: US 08/220,433
 FILING DATE: 30-MAR-1994
 APPLICATION NUMBER: US 07/914,281
 FILING DATE: 20-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Lavalleye, Jean-Paul M. P.
 REGISTRATION NUMBER: 31,451
 REFERENCE/DOCKET NUMBER: 2363-060-55
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)521-4500
 TELEFAX: (703)486-2347
 TELEFAX: 248855 JPAT UR
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 374 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-696-731-11

Query Match 88.6%; Score 31; DB 2; Length 374;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPQPG 6
 |||||
 Db 137 PTPQPG 142

RESULT 5
 US-08-319-704-6
 Sequence 6, Application US/08319704
 Patent No. 5814617
 GENERAL INFORMATION:
 APPLICANT: Hoffman, Stephen L.
 APPLICANT: Charcevit, Yudin
 APPLICANT: Hedstrom, Richard C.
 APPLICANT: Doonan, Denise L.
 TITLE OF INVENTION: Protective 17 kDa Malaria Hepatic and
 TITLE OF INVENTION: Erythrocytic Stage Immunosol and Gene
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Naval Medical R & D Command
 STREET: Bldg 1, T-12, 8901 Wisconsin Avenue
 CITY: Bethesda
 STATE: Maryland
 COUNTRY: U.S.A.
 ZIP: 20888-5606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/319,704
 FILING DATE: 07-OCT-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: A. David Spevack
 REGISTRATION NUMBER: 24,743
 REFERENCE/DOCKET NUMBER: 75,206
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 295-6759
 TELEFAX: (301) 295-1022
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 162 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-319-704-6

Query Match 82.9%; Score 29; DB 2; Length 162;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTPQPG 6
 |||||
 Db 146 PTPQPG 151

RESULT 6
 US-08-651-818A-3
 Sequence 3, Application US/08651818A
 Patent No. 5948889
 GENERAL INFORMATION:
 APPLICANT: de Boer, Piet A.J.
 APPLICANT: Hale, Cynthia A.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING
 TITLE OF INVENTION: ANTIMICROBIALS
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MEDLEN & CARROLL
 STREET: 220 Montgomery Street, Suite 2200
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/651,818A
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Carroll, Peter G.
 REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: CASE-02249
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 327 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-651-818A-3

Query Match 82.9%; Score 29; DB 2; Length 327;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPQPG 5
 |||||
 Db 123 PTPQPG 127

RESULT 7
 US-09-118-319-6
 Sequence 6, Application US/09118319
 Patent No. 6114158
 GENERAL INFORMATION:
 APPLICANT: Li, Xin-liang
 APPLICANT: Chen, Huihong
 APPLICANT: Ljungdahl, Lars G.
 TITLE OF INVENTION: Oryzomyces Cellulase Cell Protein and Coding Sequences
 FILE REFERENCE: 33-98sequence listing
 CURRENT APPLICATION NUMBER: US/09/118,319

;; CURRENT FILING DATE: 1998-07-17
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 6
;; LENGTH: 459
;; TYPE: PRT
;; ORGANISM: Ophiomyces sp. PC-2
US-09-118-319-6

Query Match 82.9%; Score 29; DB 3; Length 459;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPOG 5
11111
DB 119 PTOPOG 123

RESULT 8
US-08-749-391-2
; Sequence 2, Application US/08749391
; Patent No. 5948667
; GENERAL INFORMATION:
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Sellinger, Leonard B.
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Hu, Youji
; APPLICANT: Forsberg, Cecil W.
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: A Xylanase Obtained From an
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 3570 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80803
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,391
; FILING DATE: 13-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna M. Forbet
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 93-96
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-749-391-2

Query Match 82.9%; Score 29; DB 2; Length 485;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
11111
DB 331 PTOPOG 336

RESULT 9
US-08-676-166A-7
; Sequence 7, Application US/08676166A
; Patent No. 5955270
; GENERAL INFORMATION:
; APPLICANT: Radford, Alan
; APPLICANT: Parish, John H.
; TITLE OF INVENTION: EXPLOITATION OF THE CELLULOSE COMPLEX OF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,166A
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1321-1-002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: P. chrysosporium
US-08-676-166A-7

Query Match 82.9%; Score 29; DB 2; Length 516;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
11111
DB 474 PTOPOG 479

RESULT 10
US-09-071-101-2
; Sequence 2, Application US/09071101
; Patent No. 6013503
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Jaspers, Stephen R.
; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.101
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E. 31,648
REGISTRATION NUMBER: 97-05
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 755 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-071-101-2

Query Match 82.9%; Score 29; DB 3; Length 755;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPO 5
11111
DB 448 PTOPO 452

RESULT 11
US-09-369-618-2
Sequence 2, Application US/09369618
Patent No. 610004
GENERAL INFORMATION:
APPLICANT: Lok, Si
APPLICANT: Jasper, Stephen R.
TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
FILE REFERENCE: 37-05D2
CURRENT APPLICATION NUMBER: US/09/369,618
CURRENT FILING DATE: 1999-08-06
EARLIER FILING DATE: 1998-05-01
EARLIER APPLICATION NUMBER: US 60/044,015
EARLIER FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 755
TYPE: PRT
ORGANISM: Homo sapiens
US-09-369-618-2

Query Match 82.9%; Score 29; DB 3; Length 755;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPO 5
11111
DB 448 PTOPO 452

RESULT 12
US-08-742-753-4
Sequence 4, Application US/08742753
Patent No. 5861278
GENERAL INFORMATION:
APPLICANT: Wong, Gordon G.
APPLICANT: YAO, Kwok-Ming
TITLE OF INVENTION: HNF3-delta Compositions
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,753
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5277
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8260
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 70..2358
US-08-742-753-4

Query Match 82.9%; Score 29; DB 2; Length 763;
Best Local Similarity 83.3%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPO 6
11111
DB 115 PTOPO 120

RESULT 13
US-08-537-210A-3
Sequence 3, Application US/08537210A
Patent No. 5780300
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Fortini, Mark
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/537,210A
 FILING DATE: 29-SEP-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: MISTOCK, S. LESLIE
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-027
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-790-9090
 TELEFAX: 212-869-8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1064 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 FEATURE:
 NAME/KEY: Xen N
 LOCATION: 1150...2213
 OTHER INFORMATION: Highly conserved ankyrin repeat
 US-08-537-210A-3

Query Match
 Best Local Similarity 82.9%; Score 29; DB 1; Length 1064;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 PROPOG 6
 Db 706 PTPPOG 711

RESULT 14
 US-08-537-210A-2
 Sequence 2, Application US/08537210A
 Patent No. 5780300
 GENERAL INFORMATION:
 APPLICANT: Artavanis-Tsakonas, Spyridon
 APPLICANT: Fortini, Mark
 APPLICANT: Matsuno, Kenji
 TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY
 NUMBER OF SEQUENCES: 4
 TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10036/2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/537,210A
 FILING DATE: 29-SEP-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: MISTOCK, S. LESLIE
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-027
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-790-9090

TELEFAX: 212-869-8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1068 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 FEATURE:
 NAME/KEY: Human N1 (TAN-1)
 LOCATION: 1152...2219
 OTHER INFORMATION: Highly conserved ankyrin repeat
 US-08-537-210A-2

Query Match
 Best Local Similarity 82.9%; Score 29; DB 1; Length 1068;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 PROPOG 6
 Db 710 PTPPOG 715

RESULT 15
 US-08-264-534-32
 Sequence 32, Application US/08264534
 Patent No. 5648464
 GENERAL INFORMATION:
 APPLICANT: Artavanis-Tsakonas, Spyridon et al.
 TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
 NUMBER OF SEQUENCES: 34
 TITLE OF INVENTION: In Topotypic Proteins, And Methods Based Thereon
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/264,534
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/695,189
 FILING DATE: 03-MAY-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: MISTOCK, S. LESLIE
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 790-9090
 TELEFAX: 212 8698864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1078 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-264-534-32

Query Match
 Best Local Similarity 82.9%; Score 29; DB 1; Length 1078;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Thu Nov 16 15:44:25 2000

us-09-308-237b-1_1.rai

Page 7

Matches	5;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	1	PTQPOG	6						
		11111							
Db	407	PTPPQG	412						

Search completed: November 8, 2000, 14:37:57
Job time: 164 sec

The first part of the paper discusses the importance of maintaining accurate records of all transactions. It is essential for the business to have a clear and concise record of all income and expenses, as this will be necessary for the preparation of the tax return. The second part of the paper discusses the importance of keeping up to date with the latest tax laws and regulations. It is important to consult with a tax professional to ensure that the business is in compliance with all applicable laws. The third part of the paper discusses the importance of maintaining proper documentation for all transactions. This includes keeping receipts, invoices, and other documents that will be needed to support the tax return. The fourth part of the paper discusses the importance of keeping up to date with the latest tax software and programs. This will ensure that the business is able to accurately calculate its tax liability and prepare its tax return. The fifth part of the paper discusses the importance of keeping up to date with the latest tax forms and schedules. This will ensure that the business is able to accurately complete its tax return. The sixth part of the paper discusses the importance of keeping up to date with the latest tax rates and deductions. This will ensure that the business is able to take full advantage of all available tax benefits. The seventh part of the paper discusses the importance of keeping up to date with the latest tax credits and exemptions. This will ensure that the business is able to take full advantage of all available tax benefits. The eighth part of the paper discusses the importance of keeping up to date with the latest tax laws and regulations. This will ensure that the business is in compliance with all applicable laws. The ninth part of the paper discusses the importance of keeping up to date with the latest tax software and programs. This will ensure that the business is able to accurately calculate its tax liability and prepare its tax return. The tenth part of the paper discusses the importance of keeping up to date with the latest tax forms and schedules. This will ensure that the business is able to accurately complete its tax return. The eleventh part of the paper discusses the importance of keeping up to date with the latest tax rates and deductions. This will ensure that the business is able to take full advantage of all available tax benefits. The twelfth part of the paper discusses the importance of keeping up to date with the latest tax credits and exemptions. This will ensure that the business is able to take full advantage of all available tax benefits.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2000, 14:37:18 : Search time 36.88 Seconds
(without alignments)
21.927 Million cell updates/sec

Title: US-09-308-237B-1

Sequence: 1 prtpg 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 863609 seqs, 134777022 residues

Total number of hits satisfying chosen parameters: 863609

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA:*
1: /sgn2_6/pdata/2/paa/US056.COMB.pep:*
2: /sgn2_6/pdata/2/paa/US057.COMB.pep:*
3: /sgn2_6/pdata/2/paa/US058.COMB.pep:*
4: /sgn2_6/pdata/2/paa/US059.COMB.pep:*
5: /sgn2_6/pdata/2/paa/US060.COMB.pep:*
6: /sgn2_6/pdata/2/paa/US061.COMB.pep:*
7: /sgn2_6/pdata/2/paa/US062.COMB.pep:*
8: /sgn2_6/pdata/2/paa/US063.COMB.pep:*
9: /sgn2_6/pdata/2/paa/US064.COMB.pep:*
10: /sgn2_6/pdata/2/paa/US065.COMB.pep:*
11: /sgn2_6/pdata/2/paa/US066.COMB.pep:*
12: /sgn2_6/pdata/2/paa/US067.COMB.pep:*
13: /sgn2_6/pdata/2/paa/US068.COMB.pep:*
14: /sgn2_6/pdata/2/paa/US069.COMB.pep:*
15: /sgn2_6/pdata/2/paa/US070.COMB.pep:*
16: /sgn2_6/pdata/2/paa/US071.COMB.pep:*
17: /sgn2_6/pdata/2/paa/US072.COMB.pep:*
18: /sgn2_6/pdata/2/paa/US073.COMB.pep:*
19: /sgn2_6/pdata/2/paa/US074.COMB.pep:*
20: /sgn2_6/pdata/2/paa/US075.COMB.pep:*
21: /sgn2_6/pdata/2/paa/US076.COMB.pep:*
22: /sgn2_6/pdata/2/paa/US077.COMB.pep:*
23: /sgn2_6/pdata/2/paa/US078.COMB.pep:*
24: /sgn2_6/pdata/2/paa/US079.COMB.pep:*
25: /sgn2_6/pdata/2/paa/US080.COMB.pep:*
26: /sgn2_6/pdata/2/paa/US081.COMB.pep:*
27: /sgn2_6/pdata/2/paa/US082.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	6	US-09-308-237B-1	Sequence 1, Appl
2	35	100.0	573	PCT-US94-14643-13	Sequence 13, Appl
3	32	91.4	601	US-60-191-637-21545	Sequence 21545, A
4	32	91.4	601	US-60-191-681-16991	Sequence 16991, A
5	31	88.6	300	PCT-US99-22517-1	Sequence 1, Appl
6	31	88.6	300	US-09-162-524-1	Sequence 1, Appl

7	31	88.6	302	US-09-443-704-12	Sequence 12, Appl
8	31	88.6	374	US-08-823-489-11	Sequence 11, Appl
9	31	88.6	374	US-09-042-531-11	Sequence 11, Appl
10	30	85.7	48	US-60-162-245-2777	Sequence 2777, Ap
11	30	85.7	58	US-60-160-203-5935	Sequence 5935, Ap
12	30	85.7	58	US-60-177-571-2970	Sequence 2970, Ap
13	30	85.7	71	US-60-177-571-2820	Sequence 2820, Ap
14	30	85.7	76	US-60-160-203-4845	Sequence 4845, Ap
15	30	85.7	76	US-60-169-840-6635	Sequence 6635, Ap
16	30	85.7	76	US-60-169-841-2776	Sequence 2776, Ap
17	30	85.7	109	US-60-194-508-3365	Sequence 3365, Ap
18	30	85.7	111	US-60-197-873-22724	Sequence 22724, A
19	30	85.7	119	US-60-181-430-304	Sequence 304, App
20	30	85.7	137	US-60-189-458-157	Sequence 157, App
21	30	85.7	161	US-60-177-571-2750	Sequence 2750, Ap
22	30	85.7	163	US-60-188-631-141	Sequence 141, App
23	30	85.7	163	US-60-190-000-282	Sequence 282, App
24	30	85.7	174	US-60-189-458-158	Sequence 158, App
25	30	85.7	181	US-09-252-691-5929	Sequence 5929, App
26	30	85.7	188	US-09-417-507-34368	Sequence 34368, A
27	30	85.7	260	US-09-417-507-34912	Sequence 34912, A
28	30	85.7	264	US-09-248-736-19607	Sequence 19607, A
29	30	85.7	264	US-60-096-409-19607	Sequence 19607, A
30	30	85.7	325	US-60-212-656-344	Sequence 344, App
31	30	85.7	335	US-60-230-435-1045	Sequence 1045, App
32	30	85.7	333	US-09-417-507-40369	Sequence 40369, A
33	30	85.7	433	US-60-167-217-15227	Sequence 15227, A
34	30	85.7	433	US-60-173-464-12437	Sequence 12437, A
35	30	85.7	433	US-60-191-637-15191	Sequence 15191, A
36	30	85.7	433	US-60-191-681-12020	Sequence 12020, A
37	30	85.7	795	US-60-191-637-31777	Sequence 31777, A
38	30	85.7	858	US-60-167-217-8669	Sequence 8669, App
39	30	85.7	858	US-60-173-464-6958	Sequence 6958, App
40	30	85.7	858	US-60-191-637-8610	Sequence 8610, App
41	30	85.7	858	US-60-191-681-6693	Sequence 6693, App
42	30	85.7	2632	US-60-230-435-1528	Sequence 1528, App
43	29	82.9	40	US-09-215-026-7	Sequence 7, Appl
44	29	82.9	53	US-60-170-373-2471	Sequence 2471, Ap
45	29	82.9	65	US-60-171-494-1105	Sequence 1105, Ap

ALIGNMENTS

RESULT 1
US-09-308-237B-1
Sequence 1, Application US/09308237B
GENERAL INFORMATION:
APPLICANT: OHKI, HIENORI
APPLICANT: TOMISHIMA, MASAKI
APPLICANT: YAMADA, AKIRA
TITLE OF INVENTION: CYCLOHEXAPEPTIDES HAVING ANTIMICROBIAL ACTIVITY
FILE REFERENCE: 0018-1052-0PCT
CURRENT APPLICATION NUMBER: US/09/308,237B
PRIOR FILING DATE: 1997-05-21
PRIOR APPLICATION NUMBER: PCT/JP97/04193
PRIOR FILING DATE: 1997-11-18
PRIOR APPLICATION NUMBER: AU P03814
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 6
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1)
OTHER INFORMATION: 4-hydroxy
NAME/KEY: MOD_RES
LOCATION: (2)

OTHER INFORMATION: hydroxy, substituted phenyl
NAME/KEY: MOD_RES
LOCATION: (3)
OTHER INFORMATION: hydroxy substituted
NAME/KEY: MOD_RES
LOCATION: (4)
OTHER INFORMATION: 3-hydroxy, 4-methyl proline
NAME/KEY: MOD_RES
LOCATION: (5)
OTHER INFORMATION: hydroxy substituted
NAME/KEY: MOD_RES
LOCATION: (6)
OTHER INFORMATION: hydroxy ethyl substituted
OTHER INFORMATION: glycine at position 6 is linked to proline at
OTHER INFORMATION: position 1 by a peptide linkage
US-09-308-237b-1

Query Match 100.0%; Score 35; DB 26; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.5e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPOG 6
DB 1 PTPOG 6

RESULT 2
PCT-US94-14643-13
Sequence 13, Application PC/TUS9414643
GENERAL INFORMATION:
APPLICANT: BIOGEN, INC.
APPLICANT: CAPE, Richard L.
APPLICANT: INSEPM, (U 293)
APPLICANT: JOSSE, Nathalie
TITLE OF INVENTION: ANTI-MULLERIAN HORMONE RECEPTOR
TITLE OF INVENTION: POLYPEPTIDES AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: C/O FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14643
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/166,333
FILING DATE: 13-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,512
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: B174CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

FEATURE:
NAME/KEY: Domain
LOCATION: 18..144
OTHER INFORMATION: /note= "extracellular domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 145..170
OTHER INFORMATION: /note= "transmembrane region"
FEATURE:
NAME/KEY: Domain
LOCATION: 171..573
OTHER INFORMATION: /note= "cytoplasmic domain"
PCT-US94-14643-13

Query Match 100.0%; Score 35; DB 1; Length 573;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPOG 6
DB 369 PTPOG 374

RESULT 3
US-60-191-637-21545
Sequence 21545, Application US/60191637
GENERAL INFORMATION:
APPLICANT: Ventier, J. Craig
TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: C1000392
CURRENT APPLICATION NUMBER: US/60/191,637
CURRENT FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 42660
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21545
LENGTH: 601
TYPE: PRT
ORGANISM: DROSOPHILA
US-60-191-637-21545

Query Match 91.4%; Score 32; DB 21; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPOG 6
DB 338 PTPOG 343

RESULT 4
US-60-191-681-16991
Sequence 16991, Application US/60191681
GENERAL INFORMATION:
APPLICANT: Li, Peter, W.D.
TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
FILE REFERENCE: C1000390
CURRENT APPLICATION NUMBER: US/60/191,681
CURRENT FILING DATE: 2000-03-23
NUMBER OF SEQ ID NOS: 30973
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16991
LENGTH: 601
TYPE: PRT
ORGANISM: DROSOPHILA
US-60-191-681-16991

Query Match 91.4%; Score 32; DB 21; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTOPOG 6
1:11111
DB 338 PTOPOG 343

RESULT 5
PCT-US99-22517-1

Sequence 1, Application PC/TUS9922517
GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University, in the City o
TITLE OF INVENTION: Gene Encoding MNR2 and Uses Thereof
FILE REFERENCE: 57477 A-PCT/jpw/ml
CURRENT APPLICATION NUMBER: PCT/US99/22517
CURRENT FILING DATE: 1999-09-29
EARLIER APPLICATION NUMBER: US 09/162,524
EARLIER FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0 - beta
SEQ ID NO 1
LENGTH: 300
TYPE: PRT
ORGANISM: chick embryo
PCT-US99-22517-1

Query Match 88.6%; Score 31; DB 1; Length 300;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTOPOG 6
1:11111
DB 238 PTOPOG 243

RESULT 6

US-09-162-524-1
Sequence 1, Application US/09162524
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M.
APPLICANT: Tanabe, Yasuto
TITLE OF INVENTION: Gene Encoding MNR2 and Uses Thereof
FILE REFERENCE: 57477/jpw/ml
CURRENT APPLICATION NUMBER: US/09/162,524
CURRENT FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 300
TYPE: PRT
ORGANISM: chick embryo
US-09-162-524-1

Query Match 88.6%; Score 31; DB 16; Length 300;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTOPOG 6
1:11111
DB 238 PTOPOG 243

RESULT 7

US-09-443-704-12
Sequence 12, Application US/09443704
GENERAL INFORMATION:
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Liu, Zhan-Bin

APPLICANT: Odell, Joan
APPLICANT: Rafalski, Antoni
APPLICANT: Shi, June
APPLICANT: Wang, Zude
TITLE OF INVENTION: Plant MYB-Related Transcription Factors
FILE REFERENCE: B01280 US NA
CURRENT APPLICATION NUMBER: US/09/443,704
CURRENT FILING DATE: 1999-11-19
EARLIER APPLICATION NUMBER: 60/109,294
EARLIER FILING DATE: November 20, 1998
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Microsoft Office 97
SEQ ID NO 12
LENGTH: 302
TYPE: PRT
ORGANISM: Glycine max
US-09-443-704-12

Query Match 88.6%; Score 31; DB 18; Length 302;
Best Local Similarity 83.3%; Pred. No. 8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTOPOG 6
1:11111
DB 220 PTOPOG 225

RESULT 8

US-08-823-489-11
Sequence 11, Application US/08023489
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTO
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,489
FILING DATE: 25-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-823-489-11

Query Match 88.6% Score 31; DB 13; Length 374;
 Best Local Similarity 83.3%; Pred. No. 1e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 DB 137 PTPPOG 142

RESULT 9
 US-09-042-531-11

; Sequence 11, Application US/09042531
 ; GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.
 TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
 OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
 TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
 TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
 OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
 CORRESPONDENCE ADDRESS: 14
 ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
 STREET: P.C.
 CITY: 1755 Jefferson Davis Highway, Fourth Floor
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/042,531
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/393,246
 FILING DATE:

APPLICATION NUMBER: US/07/914,281
 FILING DATE: 30-MAR-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.
 REGISTRATION NUMBER: 31,451
 REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500
 TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 374 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-09-042-531-11

Query Match 88.6% Score 31; DB 15; Length 374;
 Best Local Similarity 83.3%; Pred. No. 1e+03;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 DB 137 PTPPOG 142

RESULT 10
 US-60-162-245-2777

; Sequence 2777, Application US/60162245
 ; GENERAL INFORMATION:
 APPLICANT: Bonazzi, Vivien
 TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
 TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
 FILE REFERENCE: C1000124
 CURRENT APPLICATION NUMBER: US/60/162,245
 CURRENT FILING DATE: 1999-10-29
 NUMBER OF SEQ ID NOS: 4178
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 2777
 LENGTH: 48
 TYPE: PRT
 ORGANISM: Human
 US-60-162-245-2777

Query Match 85.7% Score 30; DB 21; Length 48;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 DB 41 PTPPOG 46

RESULT 11
 US-60-160-203-5935

; Sequence 5935, Application US/60160203
 ; GENERAL INFORMATION:

APPLICANT: BONAZZI, VIVIEN
 TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
 NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
 USES THEREOF
 FILE REFERENCE: C1000116
 CURRENT APPLICATION NUMBER: US/60/160,203
 CURRENT FILING DATE: 1999-10-19
 NUMBER OF SEQ ID NOS: 6374
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 5935
 LENGTH: 58
 TYPE: PRT
 ORGANISM: HUMAN

FEATURE:
 NAME/KEY: VARIANT
 LOCATION: (1)-(58)
 OTHER INFORMATION: Xaa = Any Amino Acid
 US-60-160-203-5935

Query Match 85.7% Score 30; DB 21; Length 58;
 Best Local Similarity 83.3%; Pred. No. 2.1e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PTPPOG 6
 DB 33 PTPPOG 38

RESULT 12
 US-60-177-571-2970

; Sequence 2970, Application US/60177571
 ; GENERAL INFORMATION:

APPLICANT: Bonazzi, Vivien
 TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
 NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
 AND USES THEREOF
 FILE REFERENCE: C1000201
 CURRENT APPLICATION NUMBER: US/60/177,571
 CURRENT FILING DATE: 2000-01-27
 NUMBER OF SEQ ID NOS: 5082
 SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 2970
LENGTH: 58
TYPE: PRT
ORGANISM: HUMAN
US-60-177-571-2970

Query Match
Best Local Similarity 85.7%; Score 30; DB 21;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
11111
Db 16 PTDPQG 21

RESULT 13
US-60-177-571-2820
Sequence 2820, Application US/60177571

GENERAL INFORMATION:
APPLICANT: Bonazzi, Vivien
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: C1000201
CURRENT APPLICATION NUMBER: US/60/177,571
CURRENT FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 5082
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2820
LENGTH: 71
TYPE: PRT
ORGANISM: HUMAN
US-60-177-571-2820

Query Match
Best Local Similarity 85.7%; Score 30; DB 21;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
11111
Db 34 PTDPQG 39

RESULT 14
US-60-160-203-4845
Sequence 4845, Application US/60160203

GENERAL INFORMATION:
APPLICANT: BONAZZI, VIVIEN
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: C1000116
CURRENT APPLICATION NUMBER: US/60/160,203
CURRENT FILING DATE: 1999-10-19
NUMBER OF SEQ ID NOS: 6374
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4845
LENGTH: 76
TYPE: PRT
ORGANISM: HUMAN
US-60-160-203-4845

Query Match
Best Local Similarity 85.7%; Score 30; DB 21;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
11111
Db 34 PTDPQG 39

RESULT 15

US-60-169-840-6635
Sequence 6635, Application US/60169840
GENERAL INFORMATION:
APPLICANT: Bonazzi, Vivien
TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: C1000164
CURRENT APPLICATION NUMBER: US/60/169,840
CURRENT FILING DATE: 1999-12-09
NUMBER OF SEQ ID NOS: 9628
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6635
LENGTH: 76
TYPE: PRT
ORGANISM: Human
US-60-169-840-6635

Query Match
Best Local Similarity 85.7%; Score 30; DB 21;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
11111
Db 34 PTDPQG 39

Search completed: November 8, 2000, 14:39:13
Job time: 115 sec

Thu Nov 16 15:44:25 2000

us-09-308-237b-1_1.rap

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2000, 14:35:14 ; Search time 13.37 Seconds
(without alignments)
28.479 Million cell updates/sec

Title: US-09-308-237B-1
Perfect score: 35
Sequence: 1 PTPPG 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 182106 segs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR.65:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	573	2 J04335	anti-mullerian hor
2	32	91.4	792	2 H81693	outer membrane pro
3	32	91.4	792	2 B71539	probable cnp85 ana
4	31	88.6	374	2 A42270	alpha (1.3) fucosyl
5	31	88.6	557	2 S41627	mullerian-inhibiti
6	31	88.6	568	2 J05629	hypothetical prote
7	31	88.6	645	2 T27186	hypothetical prote
8	31	88.6	780	2 T00366	uricase Deimoco
9	30	85.7	298	2 C75430	cell surface glyco
10	30	85.7	573	2 A33533	pre-mRNA splicing
11	30	85.7	899	2 S12319	hypothetical prote
12	30	85.7	1160	2 T27487	perlecan homolog u
13	30	85.7	1160	2 C47648	ribonucleoside-dip
14	30	85.7	1418	2 D75281	calcium channel al
15	30	85.7	2262	2 T30890	perlecan homolog u
16	30	85.7	2277	2 B47648	hypothetical prote
17	30	85.7	2295	2 T19819	hypothetical prote
18	30	85.7	2482	2 T19819	hypothetical prote
19	30	85.7	2482	2 A47648	perlecan homolog u
20	30	85.7	3375	2 T42737	hypothetical prote
21	30	85.7	4660	2 T42087	gp330 protein prec
22	29	83.9	99	2 E69267	hypothetical prote
23	29	83.9	99	2 E69267	hypothetical prote
24	29	83.9	124	2 T41250	probable transhyr
25	29	83.9	162	1 YAK051	antigen 5.1 precu
26	29	83.9	162	2 A26768	antigen 5.1 precu
27	29	83.9	162	2 A25780	blood-stage antige
28	29	83.9	166	2 T14869	single-stranded DN
29	29	83.9	166	2 T29641	hypothetical prote

30	29	82.9	176	2 JN0632	single-stranded DN
31	29	82.9	178	1 DDEC	single-stranded DN
32	29	82.9	182	2 T41204	A-minor fibrillar p
33	29	82.9	197	2 T36584	hypothetical prote
34	29	82.9	205	2 E82032	thiamin-phosphate
35	29	82.9	205	2 A81011	carcinoembryonic a
36	29	82.9	212	2 S33324	hypothetical prote
37	29	82.9	232	2 T27396	alpha-smooth muscl
38	29	82.9	232	2 I46679	heat-shock protein
39	29	82.9	281	2 J04295	MCC2 protein precu
40	29	82.9	304	2 T18345	hypothetical prote
41	29	82.9	328	2 E64020	hypothetical prote
42	29	82.9	361	2 B59099	protein kinase pim
43	29	82.9	370	1 S53333	hypothetical prote
44	29	82.9	377	1 T47791	branched-chain ami
45	29	82.9	390	2 B69370	

ALIGNMENTS

RESULT 1
anti-mullerian hormone type II receptor precursor - human
JC4335
C:Species: Homo sapiens (man)
C:Date: 06-Dec-1995 #sequence-revision 08-Feb-1996 #text-change 04-Mar-2000
C:Accession: JC4335
R:Visser, J.A.; Mcluskey, A.; van Beers, T.; Weghuis, D.O.; van Kessel, A.G.; Grootey
Biochem. Biophys. Res. Commun. 215, 1029-1036, 1995
A:Title: Structure and chromosomal localization of the human anti-mullerian hormone
A:Reference number: JC4335; MUID:96028015
A:Accession: JC4335
A:Molecule type: mRNA
A:Residues: 1-573 <N5>
A:Cross-references: GB:J91156; NID:q1107671; PIDN:CA62593.1; PID:ej98046; PID:q11076
C:Comment: This is a receptor for anti-mullerian hormone (see PIR:WFDHM). It plays a
C:Genetics:
A:Gene: GDB:AMHR2
A:Cross-references: GDB:696210; OMIM:600956
A:Map position: 12q13-12q13
A:Insertions: 16/3; 77/3; 141/3; 167/3; 207/2; 284/2; 322/3; 380/2; 429/3; 475/2
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; hormone receptor; transmembrane protein
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-573/Product: anti-mullerian hormone type II receptor #status predicted <MAT>
F:17-141/Domain: extracellular hormone binding #status predicted <ELB>
F:142-167/Domain: transmembrane #status predicted <TM>
F:201-512/Domain: protein kinase homology <KIN>

Query Match 100.0%; Score 35; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 PTPPG 6
DB 369 PTPPG 374

RESULT 2
outer membrane protein, probable TC0512 [Imported] - Chlamydia muridarum (strain N199
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mofn
C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 11-May-2000
C:Accession: H81693
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mofn and Chlamydia pneumoniae AR39
A:Reference number: H81693
A:Accession: H81693
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-792 <TEP>
 A:Cross-references: GB:AE002320; GB:AE002160; NID:97190551; PIDN:AA639354.1; PID:9719055
 A:Experimental source: strain Nigg (Mcpn)
 C:Genetics:
 A:Gene: MC0512

Query Match 91.4%; Score 32; DB 2; Length 792;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 PROPOG 6
 Db 692 PROPOG 697

RESULT 3
 B71539
 Probable omp85 analog - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: B71539
 R:Stephens, R.S.; Kallman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis serotype D
 A:Reference number: A71570; MUID:99000809
 A:Accession: B71539
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-792 <ARN>
 A:Cross-references: GB:AE001297; GB:AE001273; NID:93328646; PIDN:AA67834.1; PID:9332865
 C:Genetics:
 A:Experimental source: serotype D, strain UW-3/Cx
 A:Gene: yaeF

Query Match 91.4%; Score 32; DB 2; Length 792;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 PROPOG 6
 Db 692 PROPOG 697

RESULT 4
 A42270
 alpha (1,3) fucosyltransferase FUT5 - human
 N:Alternate names: fucosyltransferase 5
 C:Species: Homo sapiens (man)
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 29-Sep-1999
 C:Accession: A42270; MUID:92156161
 R:Weston, B.W.; Nair, R.P.; Larsen, R.D.; Lowe, J.B. J. Biol. Chem. 267, 4152-4160, 1992
 A>Title: Isolation of a novel human alpha (1,3)fucosyltransferase gene and molecular clones encoding enzymes with distinct acceptor substrate specificities.
 A:Reference number: A42270; MUID:92156161
 A:Accession: A42270
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-374 <MES>
 A:Cross-references: GB:M81485; NID:9182490; PIDN:AA98117.1; PID:91280209
 R:Camaron, H.S.; Szczepniak, D.; Weston, B.W. J. Biol. Chem. 270, 20112-20122, 1995
 A>Title: Expression of human chromosome 19p alpha(1,3)-fucosyltransferase genes in normal cells
 A:Reference number: I39043; MUID:95378269
 A:Accession: I39046
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-374 <RES>
 A:Cross-references: EMBL:U27329; NID:9967194; PIDN:AA650188.1; PID:9967195
 A:Accession: I39047

A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-374 <RE2>
 A:Cross-references: EMBL:U27330; NID:9967196; PIDN:AA650189.1; PID:9967197
 C:Genetics:
 A:Gene: GDB:FUT5
 A:Cross-references: GDB:131644; OMIM:136835
 A:Map position: 19p13.3-19p13.3
 C:Superfamily: galactoside 3(4)-L-fucosyltransferase

Query Match 88.6%; Score 31; DB 2; Length 374;
 Best Local Similarity 83.3%; Pred. No. 73;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 PROPOG 6
 Db 137 PROPOG 142

RESULT 5
 S41627
 Probable anti-müllerian hormone receptor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
 C:Accession: S41627
 R:Baarends, W.M.; Van Helmond, M.J.L.; Post, M.; van der Schoot, P.J.C.M.; Hoogerbrug, gen. J.A. Development 120, 189-197, 1994
 A>Title: A novel member of the transmembrane serine/threonine kinase receptor family
 A:Reference number: S41627; MUID:94163972
 A:Accession: S41627
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-557 <BAA>
 A:Cross-references: EMBL:X71916; NID:9453177; PIDN:CAA50731.1; PID:9453178
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 C:Keywords: ATP
 F:199-510/Domain: protein kinase homology <KIN>

Query Match 88.6%; Score 31; DB 2; Length 557;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 PROPOG 6
 Db 367 PROPOG 372

RESULT 6
 JC5629
 müllerian-inhibiting substance type II receptor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 04-Mar-2000
 C:Accession: JC5629
 R:Michina, Y.; Tizard, R.; Deng, J.M.; Pathak, B.G.; Copeland, N.G.; Jenkins, N.A.; C. Biochem. Biophys. Res. Commun. 237, 741-746, 1997
 A>Title: Sequence, genomic organization, and chromosomal location of the mouse Muelle
 A:Reference number: JC5629; MUID:97445109
 A:Accession: JC5629
 A:Molecule type: DNA
 A:Residues: 1-568 <MTS>
 C:Comment: This receptor is involved in the sexual differentiation.
 C:Genetics:
 A:Gene: Amhr
 A:Map position: 15
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 F:197-508/Domain: protein kinase homology <KIN>

Query Match 88.6%; Score 31; DB 2; Length 568;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PTPPOG 6
 |||||
 Db 365 PTPPOG 370

RESULT 7

hypothetical protein Y54G9A.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27186

R:Smyle, R.
 Submitted to the EMBL Data Library, October 1998

A:Reference number: 220324

A:Accession: T27186

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-645 <WIL>
 A:Cross-references: EMBL:AL032648; PIDN:CAA21699.1; GSPDB:GN00020; CESP:Y54G9A.3

A:Experimental source: clone Y54G9A

C:Genetics:

A:Gene: CESP:Y54G9A.3

A:Map position: 2

A:Introns: 56/3; 100/2; 148/1; 411/2; 541/2; 575/3

Query Match 88.6%; Score 31; DB 2; Length 645;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPPOG 6
 |||||
 Db 472 PTPPOG 477

RESULT 8

hypothetical protein KIAA0669 - human

C:Species: *Homo sapiens* (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Oct-1999

C:Accession: T00366

R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
 DNA Res. 5, 169-176, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete

A:Reference number: Z14142; MUID:98403880

A:Accession: T00366

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-780 <IS>

A:Cross-references: EMBL:AB014569; NID:d1204324; PIDN:BA31644.1; PID:d1032605

A:Experimental source: Brain; clone HK02346

C:Genetics:

A:Note: KIAA0669

Query Match 88.6%; Score 31; DB 2; Length 780;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPPOG 6
 |||||
 Db 311 PTPPOG 316

RESULT 9
 C75430
 urticase - *Deinococcus radiodurans* (strain R1)

C:Species: *Deinococcus radiodurans*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: C75430

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

, M.; Shen, M.; Yamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250; MUID:20036896

A:Accession: C75430

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <WHI>
 A:Cross-references: GB:AE001965; GB:AE000513; NID:96458892; PIDN:MAF10733.1; PID:9645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRI160

A:Map position: 1

C:Superfamily: urate oxidase

Query Match 85.7%; Score 30; DB 2; Length 298;
 Best Local Similarity 83.3%; Pred. No. 91;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTPPOG 6
 |||||
 Db 125 PTPPOG 130

RESULT 10

A33533
 cell surface glycoprotein precursor - mouse

C:Species: *Mus musculus* (house mouse)

C:Date: 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change 11-Jan-2000

C:Accession: A33533

R:Dougherty, G.J.; Kay, R.J.; Humphries, R.K.

J. Biol. Chem. 264, 6509-6514, 1989

A:Title: Molecular cloning of 114/110, a cell surface antigen containing highly conse

lines.

A:Reference number: A33533; MUID:89197960

A:Accession: A33533

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-573 <DOU>
 A:Cross-references: GB:J04634; NID:9191943; PIDN:MAA37239.1; PID:9309106

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: glycoprotein; transmembrane protein

F:423-466/Domain: EGF homology <DGR>

Query Match 85.7%; Score 30; DB 2; Length 573;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTPPOG 6
 |||||
 Db 213 PTPPOG 218

RESULT 11

S12319
 pre-mRNA splicing factor PRP6 - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein YBR0508; protein YBR055c

C:Species: *Saccharomyces cerevisiae*

C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 17-Mar-1999

C:Accession: S12319; S45913; S49510; S55854

R:Legrain, P.; Choulika, A.

EMBO J. 9, 2775-2781, 1990

A:Title: The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cy

A:Reference number: S12319; MUID:90360988

A:Accession: S12319

A:Molecule type: DNA

A:Residues: 1899 <LEG>

A:Cross-references: EMBL:X33465; NID:94238; PID:94239

R:Aljinovic, G.; Pohll, F.M.; Pohll, T.M.
 Submitted to the Protein Sequence Database, August 1994

A:Reference number: S45906
 A:Accession: S45913
 A:Molecule type: DNA
 A:Residues: 1-899 <AL2>
 A:Cross-references: EMBL:Z35924; NID:9536290; PID:9536291; MIPS:YBR055C
 A:Experimental source: strain S288C
 R:Aljovic, G.
 A:Title: Sequence and analysis of 24 kb on chromosome II of *Saccharomyces cerevisiae*
 A:Description: Sequence and analysis of 24 kb on chromosome II of *Saccharomyces cerevisiae*
 A:Reference number: S49503
 A:Accession: S49510
 A:Molecule type: DNA
 A:Residues: 1-899 <AL2>
 A:Cross-references: EMBL:Z46260; NID:9559942; PID:9559951
 A:Experimental source: strain S288C
 R:Aljovic, G.; Pohl, T.M.
 A:Title: Sequence and analysis of 24 kb on chromosome II of *Saccharomyces cerevisiae*
 A:Reference number: S55854
 A:Accession: S55854
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-899 <AL2>
 A:Cross-references: EMBL:Z46260; NID:9559942; PID:9559951
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
 C:Genetics:
 A:Gene: SGD:PRP6
 A:Cross-references: SGD:S0000259; MIPS:YBR055C
 A:Map position: 2R
 C:Keywords: nucleus; RNA binding; zinc finger

Query Match 85.7%; Score 30; DB 2; Length 899;
 Best Local Similarity 83.3%; Pred. No. 2.9e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 PRP06 6
 Db 238 PRP06 243

RESULT 12
 T27487
 hypothetical protein ZC101.2b - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T27487
 R:Percy, C.
 A:Submitted to the EMBL Data Library, March 1997
 A:Reference number: Z20375
 A:Accession: T27487
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1160 <M11>
 A:Cross-references: EMBL:Z93395; PIDN:CA07704.1; GSPDB:GM00020; CESP:ZC101.2b
 A:Experimental source: clone ZC101
 C:Genetics:
 A:Gene: CESP:ZC101.2b
 A:Map position: 2
 A:Introns: 32/1; 134/1; 225/1; 335/2; 450/3; 739/3; 830/3; 860/2; 1064/2

Query Match 85.7%; Score 30; DB 2; Length 1160;
 Best Local Similarity 83.3%; Pred. No. 3.8e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 PRP06 6
 Db 712 PRP06 717

RESULT 13
 C47648

perlecan homolog unc-52, form 3 precursor - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 16-Jul-1999
 C:Accession: C47648
 R:Rogalski, T.M.; Williams, B.D.; Mullen, G.P.; Moerman, D.G.
 A:Title: Products of the unc-52 gene in *Caenorhabditis elegans* are homologous to the
 A:Reference number: A47648; MUID:93339574
 A:Accession: C47648
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1160 <ROG>
 A:Cross-references: GB:L13458
 C:Superfamily: LDL receptor ligand-binding repeat homology; laminin-type EGF-like hom
 F:149-183/Domain: LDL receptor ligand-binding repeat homology <LDL1>
 F:190-224/Domain: LDL receptor ligand-binding repeat homology <LDL2>
 F:233-268/Domain: LDL receptor ligand-binding repeat homology <LDL3>
 F:955-1002/Domain: laminin-type EGF-like homology <LEC1>
 F:1011-1058/Domain: laminin-type EGF-like homology <LEC2>

Query Match 85.7%; Score 30; DB 2; Length 1160;
 Best Local Similarity 83.3%; Pred. No. 3.8e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 PRP06 6
 Db 712 PRP06 717

RESULT 14
 D75281
 ribonucleoside-diphosphate reductase-related protein - *Deinococcus radiodurans* (strai
 C:Species: *Deinococcus radiodurans*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: D75281
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* RI.
 A:Reference number: A75250; MUID:20036896
 A:Accession: D75281
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1418 <M11>
 A:Cross-references: GB:AE002068; GB:AE000513; NID:96460186; PIDN:AAFI1919.1; PID:9646
 A:Experimental source: strain RI
 C:Genetics:
 A:Gene: DR2374
 A:Map position: 1

Query Match 85.7%; Score 30; DB 2; Length 1418;
 Best Local Similarity 83.3%; Pred. No. 4.6e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 PRP06 6
 Db 233 PRP06 238

RESULT 15
 T30890
 calcium channel alpha1-chain - sea squirt (*Halocynthia roretzi*)
 N:Alternate names: ascidian calcium channel alpha1-subunit
 C:Species: *Halocynthia roretzi*
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T30890
 R:Okamura, Y.; Okagaki, R.
 A:Submitted to the EMBL Data Library, May 1998
 A:Description: T3Cal, ascidian calcium channel alpha1-subunit.
 A:Reference number: Z20924

A:Accession: T30890
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2262 <OKA>
 A:Cross-references: EMBL:AB013604; FIDN:BAA34927.1
 A:Experimental source: Young tadpole larvae
 C:Genetics:
 A:Gene: TUCa1
 C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain
 C:Keywords: calcium binding; calcium channel

Query Match 85.7%; Score 30; DB 2; Length 2262;
 Best Local Similarity 83.3%; Pred. No. 7.5e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
 |||||
 Db 1839 PTOPOG 1844

Search completed: November 8, 2000, 14:38:34
 Job time: 200 sec

GenCode version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2000, 14:38:03 ; Search time 8.45 Seconds
(without alignments)
22.685 Million cell updates/sec

Title: US-09-308-237b-1

Perfect score: 35

Sequence: 1 PROPOG 6

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	35	100.0	573	1 AMH2_HUMAN	Q16671 homo sapien
2	31	88.6	374	1 FUT5_HUMAN	Q11128 homo sapien
3	31	88.6	374	1 FUT5_PANTR	P56433 pan troglod
4	31	88.6	557	1 AMH2_RAT	O62893 rattus norv
5	30	85.7	398	1 TBX1_HUMAN	O43435 homo sapien
6	30	85.7	543	1 TR49_CAEEL	Q09392 caenorhabd
7	30	85.7	573	1 C114_MOUSE	P19467 mus musculu
8	30	85.7	899	1 PR06_YEAST	P19735 saccharomyc
9	30	85.7	1394	1 CNG4_BOVIN	Q28181 bos taurus
10	30	85.7	2481	1 UN52_CAEEL	O06561 caenorhabd
11	30	85.7	4660	1 LRP2_RAT	P98158 rattus norv
12	29	82.9	40	1 MPAL_USTMA	P31962 ustiliago ma
13	29	82.9	162	1 CRA1_PLAFA	P04923 plasmodium
14	29	82.9	162	1 EXP1_PLAFA	P04926 plasmodium
15	29	82.9	175	1 SSB_SERMA	P25762 serratia ma
16	29	82.9	177	1 SSB_ECOLI	P02339 escherichia
17	29	82.9	281	1 HS32_DICDI	P54658 dictyostell
18	29	82.9	304	1 P32_MYCGA	O49378 mycoplasma
19	29	82.9	328	1 ZIFA_HAEIN	P44113 haemophilus
20	29	82.9	516	1 P32_MYCGA	P13860 phanerocha
21	29	82.9	568	1 YG54_YEAST	P53309 saccharomyc
22	29	82.9	578	1 YK08_CAEEL	O20471 caenorhabd
23	29	82.9	657	1 KNOB_PLAFA	P06719 plasmodium
24	29	82.9	708	1 GBF_DICDI	P36417 dictyostell
25	29	82.9	763	1 FXM1_HUMAN	Q08050 h forkhead
26	29	82.9	1174	1 KPCI_COCH	O42632 cochllobolu
27	29	82.9	1189	1 YJH6_YEAST	P47035 saccharomyc
28	29	82.9	1460	1 CAL1_CANFA	O94317 canis famli
29	29	82.9	2414	1 NOTC_HUMAN	P31000 homo sapien
30	29	82.9	2437	1 NOTC_BRARE	P46530 brachydantio
31	29	82.9	2444	1 NOTC_HUMAN	P46531 homo sapien
32	29	82.9	2524	1 NOTC_XENLA	P21783 xenopus lae
33	29	82.9	2531	1 NTC1_MOUSE	O01705 mus musculu

34	29	82.9	2531	1 NTC1_RAT	O07008 rattus norv
35	28	80.0	143	1 TS22_MOUSE	O00992 mus musculu
36	28	80.0	144	1 TS22_HUMAN	Q15714 homo sapien
37	28	80.0	246	1 YP73_MTCU	Q50646 mycobacteri
38	28	80.0	308	1 Y775_METTR	O26869 methanobact
39	28	80.0	359	1 OMPA_SERMA	P04845 serratia ma
40	28	80.0	366	1 CAS4_EPHMU	P16503 ephydalia m
41	28	80.0	463	1 D5DR_FUGRU	P53454 figu rubrip
42	28	80.0	469	1 COG1_BOVIN	P28053 bos taurus
43	28	80.0	582	1 SYN2_HUMAN	O92777 homo sapien
44	28	80.0	586	1 SYN2_RAT	O63537 rattus norv
45	28	80.0	657	1 HUTH_HUMAN	P42357 homo sapien

ALIGNMENTS

RESULT 1
ID AMH2_HUMAN STANDARD: PRT: 573 AA.
AC 016671; Q13762;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ANTI-MULLERIAN HORMONE TYPE II RECEPTOR PRECURSOR (EC 2.7.1.-) (AMH
TYPE II RECEPTOR) (MIS TYPE II RECEPTOR) (MIRII) (MRLII).
GN AMH2 OR AMHR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP MEDLINE: 96083584.
RA Tizard R., Carre-Pusebe D., Belville C., Trageion L., Tonkin C.,
RA Nelson J., McAniff M., Bidart J.-M., Lababidi A., Joso N.,
RA Cate R.L., Picard J.-Y.;
RT "Insensitivity to anti-mullerian hormone due to a mutation in the
human anti-mullerian hormone receptor.";
RL Nat. Genet. 11:382-388(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96028015.
RA Visser J.A., Meluskey A., van Beers T., Weghuis D.O., van Kessel A.G.,
RA Grootegeed J.A., Thermen A.P.N.;
RT "Structure and chromosomal localization of the human anti-mullerian
hormone type II receptor gene.";
RL Biochem. Biophys. Res. Commun. 215:1029-1036(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTS;
RA Maslakov P.T., MacLaughlin D.T., Maheswaran S., Teixeira J.M.,
RA Fuller A.F., Shah P.C., Kehas D.J., Kennelly M.K., Dombkowski D.M.,
RA Ha T.V., Pfeiffer F.I., Donahoe P.K.;
RT "Human ovarian cancer, cell lines and primary ascites cells, express
the human MIS type II, blind, and are responsive to mis.";
RL Clin. Cancer Res. 0:0-0(1999).
CC -I- FUNCTION: RECEPTOR FOR ANTI-MULLERIAN HORMONE.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -I- DISEASE: DEFECTS IN AMH2 ARE THE CAUSE OF PERSISTENT MULLERIAN
DOCT SYNDROME TYPE II (PMDS-2); A FORM OF MALE
PSEUDOHERMAPHRODITISM CHARACTERIZED BY A FAILURE OF MULLERIAN
DOCT REGRESSION IN OTHERWISE NORMAL MALES.
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TGFβ RECEPTOR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

```

CC EMBL: X91156; CA62593.1; JOINED.
DR EMBL: X91157; CA62593.1; JOINED.
DR EMBL: X91158; CA62593.1; JOINED.
DR EMBL: X91159; CA62593.1; JOINED.
DR EMBL: X91160; CA62593.1; JOINED.
DR EMBL: X91161; CA62593.1; JOINED.
DR EMBL: X91162; CA62593.1; JOINED.
DR EMBL: X91163; CA62593.1; JOINED.
DR EMBL: X91164; CA62593.1; JOINED.
DR EMBL: X91165; CA62593.1; JOINED.
DR EMBL: X89013; CA61418.1; -.
DR EMBL: U29700; AAC50328.1; -.
DR EMBL: AF172932; AAD48497.1; -.
DR MIN: 600956; -.
DR INTERPRO: IPR000472; -.
DR INTERPRO: IPR000719; -.
DR PFAM: PF00069; PKinase; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KM Receptor: Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Transmembrane; Glycoprotein; Signal; Pseudotransmembrane;
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 573 ANTI-MULLERIAN HORMONE TYPE II RECEPTOR.
FT DOMAIN 18 149 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 150 170 POTENTIAL.
FT DOMAIN 171 573 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 203 518 PROTEIN KINASE.
FT NP_BIND 209 217 ATP (BY SIMILARITY).
FT BINDING 230 230 ATP (BY SIMILARITY).
FT ACT_SITE 333 333 BY SIMILARITY.
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 161 161 L -> V (IN CA62593).
SQ SEQUENCE 573 AA; 62749 MW; 1347C10C2942EDBA CRC64;

Query Match 100.0%; Score 35; DB 1; Length 573;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPPOG 6
Db 369 PTPPOG 374

RESULT 2
FUT5_HUMAN STANDARD; PRT; 374 AA.
ID FUT5_HUMAN
AC 011128;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ALPHA-(1,3)-FUCOSYLTRANSFERASE (EC 2.4.1.65) (GALACTOSIDE 3-L-
DE FUCOSYLTRANSFERASE) (FUCOSYLTRANSFERASE 5) (FUCT-V).
GN FUT5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PERIPHERAL BLOOD LEUKOCYTES;
RX MEDLINE; 92156161.
RA Weston B.W., Malt R.P., Larsen R.D., Lowe J.B.;
RT "Isolation of a novel human alpha (1,3)fucosyltransferase gene and
RT molecular comparison to the human Lewis blood group alpha
RT (1,3/1,4)fucosyltransferase gene. Systemic, homologous, nonallelic
RT genes encoding enzymes with distinct acceptor substrate
RT specificities";
RL J. Biol. Chem. 267:4152-4160(1992).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-COLON, KIDNEY, AND LIVER;
RX MEDLINE; 95378269.
RA Cameron H.S., Szczepaniak D., Weston B.W.;
RT "Expression of human chromosome 19p alpha(1,3)-fucosyltransferase
RT genes in normal tissues. Alternative splicing, polyadenylation, and
RT isoforms.";
RL J. Biol. Chem. 270:20112-20122(1995).
CC -1- FUNCTION: MAY CATALYSE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
CC THE EXPRESSION OF YIM-2, LEWIS X/SSA-1 AND SIALYL LEWIS X
CC ANTIGENS.
CC -1- CATALYTIC ACTIVITY: GDP-L-FUCOSE + 1,3-BETA-D-GALACTOSYL-
CC N-ACETYL-D-GLUCOSAMINYL-R = GDP + 1,3-BETA-D-GALACTOSYL-
CC (ALPHA-1,4-L-FUCOSYL)-N-ACETYL-D-GLUCOSAMINYL-R.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
CC FORM IN TRANS CISTERNAE OF GOLGI.
CC -1- TISSUE SPECIFICITY: LIVER, COLON AND TESTIS AND TRACE AMOUNTS IN
CC T-CELLS AND BRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: M81485; AAA98117.1; -.
DR EMBL: U27329; AAC50188.1; -.
DR EMBL: U27330; AAC50189.1; -.
DR MIN: 136835; -.
DR INTERPRO: IPR001503; -.
DR PFAM: PF00852; Fucosyltransferase; 1.
KW Signal-anchor; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack.
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 16 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT 35 374 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 374 AA; 43008 MW; B825281521B57939 CRC64;

Query Match 88.6%; Score 31; DB 1; Length 374;
Best Local Similarity 83.3%; Pred. No. 38;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTPPOG 6
Db 137 PTPPOG 142

RESULT 3
FUT5_PANTR STANDARD; PRT; 374 AA.
ID FUT5_PANTR
AC P56433;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALPHA-(1,3)-FUCOSYLTRANSFERASE (EC 2.4.1.65) (GALACTOSIDE 3-L-
DE FUCOSYLTRANSFERASE) (FUCOSYLTRANSFERASE 5) (FUCT-V).
GN FUT5.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98037800.
RA Costache W., Apoll P.-A., Calileau A., Elmgren A., Larson G.,

```

RA Henry S., Blancher A., Iordachescu D., Ortol R., Molligone R.;
 RT "Evolution of glycosyltransferase genes in vertebrates.";
 RL J. Biol. Chem. 272:29721-29728(1997).
 CC -1- FUNCTION: MAY CATALYZE ALPHA-1,3 GLYCOSIDIC LINKAGES INVOLVED IN
 CC THE EXPRESSION OF VIM-2, LEWIS X/SSA-1 AND SIALYL LEWIS X
 CC ANTIGENS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: GDP-L-FUCOSE + 1,3-BETA-D-GALACTOSYL-
 CC N-ACETYL-D-GLUCOSAMINY-L-R = GDP + 1,3-BETA-D-GALACTOSYL-
 CC (ALPHA-1,4-L-FUCOSYL)-N-ACETYL-D-GLUCOSAMINY-L-R.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 CC FORM IN TRANS CISTERNAE OF GOLGI.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y14034; CA74361.1; -.
 CC DR INTERPRO: IPR001503; -.
 CC PRAM: PF00852; Flocosyl.transf. 1.
 CC KW Transferase: Glycosyltransferase; Glycoprotein; Transmembrane;
 CC Signal-anchor; Gcgl stack.
 CC FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 16 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC FT DOMAIN 35 374 LUMENAL, CATALYTIC (POTENTIAL).
 CC FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 374 AA; 43034 MW; 3F35DEB8B8F1A2E2 CnC64;

Query Match 88.6%; Score 31; DB 1; Length 374;
 Best Local Similarity 83.3%; Pred. No. 38;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
 DB 137 PTRPOG 142

RESULT 4
 AMH2_RAT STANDARD: PRT; 557 AA.
 AC 062893: 063045: (Q9R0AT);
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DE ANTI-MULLERIAN HORMONE TYPE II RECEPTOR PRECURSOR (EC 2.7.1.-) (AMH
 DE TYPE II RECEPTOR (MIS TYPE II RECEPTOR) (MISRII) (MII) (CI14).
 GN AMH2.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 9610713;
 RA Teixeira J., He W.W., Shah P.C., Morikawa N., Lee M.M., Catlin E.A.,
 RA Hudson P.L., Wing J., MacLaughlin D.T., Donahoe P.K.;
 RT "Developmental expression of a candidate mullerian inhibiting
 RT substance type II receptor.";
 RL Endocrinology 137:160-165(1996).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE-TESTIS
 RX MEDLINE: 94163972;
 RA Baarends W.M., Van Helmond M.J.L., Post M., Van der Schoot P.J.C.M.,
 RA Hoogerbrugge J.W., de Winter J.P., Uilenbroek J.T.J., Karels B.,

RA Wilming L.G., Meijers J.H.C., Themmen A.P.N., Grootegoed A.J.;
 RT "A novel member of the transmembrane serine/threonine kinase receptor
 RT family is specifically expressed in the gonads and in mesenchymal
 RT cells adjacent to the mullerian duct.";
 RL Development 120:189-197(1994).
 RN [3]
 RN SEQUENCE OF 1-16 FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY;
 RA Teixeira J., Kehas P.J., Antun R., Donahoe P.K.;
 RT "Transcriptional regulation of the rat mullerian inhibiting substance
 RT type II receptor in rodent Leydig cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13831-13838(1999).
 CC -1- FUNCTION: RECEPTOR FOR ANTI-MULLERIAN HORMONE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC SURROUNDING THE MULLERIAN DUCT AT EMBRYONIC DAYS 14, 15, AND 16
 CC AND IN TUBULAR AND FOLLICULAR STRUCTURES OF THE FETAL GONADS.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC TGBR RECEPTOR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U42427; AAC52343.1; -.
 CC DR EMBL: X71916; CA50731.1; -.
 CC DR EMBL: AF092445; AAC64138.1; -.
 CC DR INTERPRO: IPR000472; -.
 CC DR INTERPRO: IPR000719; -.
 CC DR PRAM: PF00069; kinase; 1.
 CC DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE NEG.
 CC DR PROSITE: PS00108; PROTEIN KINASE-ST; FALSE NEG.
 CC DR PROSITE: PS50011; PROTEIN KINASE-DOM; 1.
 CC KW Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding;
 CC Transmembrane; Glycoprotein; Signal.
 CC FT SIGNAL 1 17 POTENTIAL.
 CC FT CHAIN 18 557 ANTI-MULLERIAN HORMONE TYPE II RECEPTOR.
 CC FT DOMAIN 18 144 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 145 165 POTENTIAL.
 CC FT DOMAIN 166 557 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 201 511 PROTEIN KINASE.
 CC FT NP_BIND 207 215 ATP (BY SIMILARITY).
 CC FT BINDING 228 228 ATP (BY SIMILARITY).
 CC FT ACT_SITE 331 331 BY SIMILARITY.
 CC FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CONFLICT 527 527 C -> Y (IN REF. 2).
 CC SQ SEQUENCE 557 AA; 59748 MW; 8EDBE9C0C32EBDD5 CnC64;

Query Match 88.6%; Score 31; DB 1; Length 557;
 Best Local Similarity 83.3%; Pred. No. 58;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PROPOG 6
 DB 367 PTRPOG 372

RESULT 5
 TBX1_HUMAN STANDARD: PRT; 398 AA.
 AC 043435: 043436: (PRT);
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-MAY-2000 (Rel. 39, Last annotation update)
 DE TBX1 PROTEIN (T-BOX PROTEIN 1) (TESTIS-SPECIFIC T-BOX PROTEIN).
 GN TBX1.
 OS Homo sapiens (Human).

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE, AND TESTIS;
RX MEDLINE: 97422603.
RA Chigffo C., Garvey N., Gong W., Roe B., Zhang G., Silver L.,
RA Emanuel B.S., Budarf M.L.;
RT "Isolation and characterization of a gene from the Digeorge
RT chromosomal region homologous to the mouse Tpx1 gene.";
RL Genomics 43:267-277(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF012130; AAB94018.1; -
DR EMBL: AF012131; AAB94019.1; -
DR MIM: 602054; -
DR INTERPRO: IPR001699; -
DR PFAM: PF00907; T-box; 1.
DR PRINTS: PR00937; T-box.
DR PROSITE: PS01283; TBOX_1; 1.
DR PROSITE: PS01264; TBOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW Alternative splicing.
FT DOMAIN 43 48 POLY-PRO.
FT DOMAIN 54 57 POLY-ALA.
FT DOMAIN 61 67 POLY-PRO.
FT DOMAIN 94 99 POLY-ALA.
FT DVA_BIND 119 297 T-BOX.
FT VARSPIC 338 398 GHVLDKDKVAENSRNTPREVELLRDAGGVNLTGPPEAE
FT COPENTGGLVAGPTADRLC -> LYESGGDQPLGLDVL
FT KPPSKSESLRPKHCNDT (IN ISOFORM B).
FT SQ SQUINQUE 398 AA; 43133 MW; FAF0F3FA0CC6176 CRC64;

```

```

Query Match      85.7%; Score 30; DB 1; Length 398;
Best Local Similarity 83.3%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 PTPPG 6
Db 328 PTPPG 333

```

```

RESULT 6
YR49_CAEELI. STANDARD; PRT; 545 AA.
AC 009392;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHELICAL 61.9 KDA PROTEIN F47D12.9 IN CHROMOSOME III.
GN F47D12.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
RN [1]
RP SOURCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Tatch A.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U22831; AAB64320.1; -
DR WORMPEP: F47D12.9; CE01952.
KW Hypothetical protein.
SQ SEQUENCE 545 AA; 61946 MW; 50F0BD70CE9A343C CRC64;

```

```

Query Match      85.7%; Score 30; DB 1; Length 545;
Best Local Similarity 83.3%; Pred. No. 89;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 PTPPG 6
Db 77 PTPPG 82

```

```

RESULT 7
C114_MOUSE
ID C114_MOUSE STANDARD; PRT; 573 AA.
AC P19467;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE CELL SURFACE ANTIGEN 114/A10 PRECURSOR.
CN LY64.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89197960.
RA Dougherty G.J., Kay R.J., Humphries R.K.;
RT "Molecular cloning of 114/A10, a cell surface antigen containing
RT highly conserved repeated elements, which is expressed by murine
RT hemopoietic progenitor cells and Interleukin-3-dependent cell
RT lines."
RT J. Biol. Chem. 264:6509-6514(1989).
CC -1- FUNCTION: MAY HAVE A POSITIVE REGULATORY ROLE IN THE CELLULAR
CC RESPONSE TO IL-3.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- PTM: THE SER/THR-RICH TANDEN REPEATS MAY SERVE AS SITES OF
CC EXTENSIVE GLYCOSYLATION.
CC -1- PTM: THE MOTIF SER-GLY CONSERVED IN ALL 8 SER/THR-RICH REPEATS MAY
CC SERVE AS THE SITE OF ATTACHMENT OF GLYCOSAMINOGLYCAN SIDE CHAINS.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: J04634; AAA37239.1; -
DR PIR: A33533; A33533.
DR MGD: MGI:103190; LY64.
DR INTERPRO: IPR000082; -
DR INTERPRO: IPR000561; -
DR PFAM: PF01390; SEA; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 2.
KW Glycoprotein; Signal; Antigen; EGF-like domain; Repeat; Transmembrane.
FT CHAIN 1 573 POTENTIAL.
FT DOMAIN 18 480 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 481 508 POTENTIAL.

```



```

FT DOMAIN 509 573 CYTOPLASMIC (POTENTIAL)
FT REPEAT 17 230 8 X TANDEM REPEATS, SER/THR-RICH.
FT REPEAT 17 43 1.
FT REPEAT 44 70 2.
FT REPEAT 71 97 3.
FT REPEAT 98 124 4.
FT REPEAT 125 151 5.
FT REPEAT 152 178 6.
FT REPEAT 179 205 7 (NEAR PERFECT).
FT REPEAT 206 230 8 (APPROXIMATE).
FT DOMAIN 233 273 EGF-LIKE 1.
FT DOMAIN 385 425 EGF-LIKE 2.
FT DOMAIN 425 467 EGF-LIKE 3.
FT DISULFID 237 248 BY SIMILARITY.
FT DISULFID 242 257 BY SIMILARITY.
FT DISULFID 259 272 BY SIMILARITY.
FT DISULFID 389 402 BY SIMILARITY.
FT DISULFID 394 408 BY SIMILARITY.
FT DISULFID 410 424 BY SIMILARITY.
FT DISULFID 429 441 BY SIMILARITY.
FT DISULFID 433 451 BY SIMILARITY.
FT CARBOHD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 573 AA; 56701 MW; 1154C4F04E2D58A9 CRC64;

```

```

Query Match 85.7%; Score 30; DB 1; Length 573;
Best Local Similarity 83.3%; Pred. No. 94;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 PTPPOG 6
Db 213 PTPPOG 218

```

```

RESULT 8
PROG_YEAST STANDARD; PRT; 899 AA.
ID P19735;
AC 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE PRE-MRNA SPLICING FACTOR PRP6.
DE PRP6 OR RNA6 OR YBR055C OR YBR0508.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-S288C / GRP88;
RA MEDLINE; 90360988.
RA Legrain P., Choulika A.;
RT "The molecular characterization of PRP6 and PRP9
RT a new cysteine/histidine motif common to several
RT yeast splicing factors."
RL EMBJ J. 9:2775-2781(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC SPRAIN-S288C.
RA MEDLINE; 95321020.
RA Aljinovic G., Pohl T.M.;
RT "Sequence and analysis of 24 kb on chromosome II of Saccharomyces
RT cerevisiae."
RL Yeast 11:475-479(1995).
CC -1- FUNCTION: INVOLVED IN PRE-MRNA SPLICING.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: TO SOME OTHER PRE-MRNA SPLICING FACTORS.
CC -1- SIMILARITY: CONTAINS 10 TPR DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL; X53465; CAA37559.1; -
DR EMBL; 235924; CAA84998.1; -
DR EMBL; 246260; CAA86398.1; -
DR PIR; S12319; S12319.
DR SGD; S0000259; PRP6.
KW mRNA processing; mRNA splicing; TPR domain; Repeat; Nuclear protein.
FT REPEAT 225 258 TPR.
FT REPEAT 259 290 TPR.
FT REPEAT 291 321 TPR.
FT REPEAT 322 352 TPR.
FT REPEAT 353 384 TPR.
FT REPEAT 385 414 TPR.
FT REPEAT 487 528 TPR.
FT REPEAT 686 717 TPR.
FT REPEAT 755 786 TPR.
FT REPEAT 855 886 TPR.
SQ SEQUENCE 899 AA; 104228 MW; 3E87F8AF63EAC41 CRC64;

```

```

Query Match 85.7%; Score 30; DB 1; Length 899;
Best Local Similarity 83.3%; Pred. No. 15e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 PTPPOG 6
Db 238 PTPPOG 243

```

```

RESULT 9
CNG4_BOVIN STANDARD; PRT; 1394 AA.
ID Q28181; Q28082; Q03861;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 240 KDA PROTEIN OF ROD PHOTORECEPTOR CNG-CHANNEL [CONTAINS: GLUTAMIC
DE ACID-RICH PROTEIN (GARP): CYCLIC-NUCLEOTIDE-GATED CATION CHANNEL 4
DE (CNG CHANNEL 4) (CNG-4) (CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
DE MODULATORY SUBUNIT)].
GN CNG4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC MEDLINE; 96009859.
RA Koerschen H.G., Tilling M., Seifert R., Sesti F., Williams A.,
RA Goetzes S., Colville C., Mueller F., Dose A., Godde M., Molday L.,
RA Kaupp U.B., Molday R.S.;
RT "A 240 kDa protein represents the complete beta subunit of the cyclic
RT nucleotide-gated channel from rod photoreceptor."
RL Neuron 15:627-636(1995).
RN [2]
RP SEQUENCE OF 454-1394 FROM N.A.
RC TISSUE-TESTIS.
RA MEDLINE; 9618098.
RA Biel M., Zong X., Ludwig A., Sauter A., Hofmann F.;
RT "Molecular cloning and expression of the Modulatory subunit of the
RT cyclic nucleotide-gated cation channel."
RL J. Biol. Chem. 271:6349-6355(1996).
RN [3]
RP SEQUENCE OF 1-590 FROM N.A.
RC TISSUE-RETINA.
RA Sugimoto Y., Yatsunami K., Tsujimoto M., Khorana H.G., Ichikawa A.;
RL Submitted (XXS-1991) to the EMBL/Genbank/DBJ databases.
CC -1- SUBUNIT: FORMS FUNCTIONAL HETEROLOGIC CHANNELS WITH CNG3.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

```


FT DOMAIN 730 920 LAMININ DOMAIN IV 2.
 FT DOMAIN 921 953 LAMININ EGF-LIKE 4 (C-TERMINAL).
 FT DOMAIN 954 1003 LAMININ EGF-LIKE 5.
 FT DOMAIN 1010 1059 LAMININ EGF-LIKE 6.
 FT DOMAIN 1060 1110 LAMININ EGF-LIKE 7.
 FT DOMAIN 1133 1216 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 1227 1311 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 1318 1403 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 1415 1499 IG-LIKE C2-TYPE DOMAIN 6.
 FT DOMAIN 1507 1592 IG-LIKE C2-TYPE DOMAIN 7.
 FT DOMAIN 1598 1682 IG-LIKE C2-TYPE DOMAIN 8.
 FT DOMAIN 1699 1785 IG-LIKE C2-TYPE DOMAIN 9.
 FT DOMAIN 1794 1878 IG-LIKE C2-TYPE DOMAIN 10.
 FT DOMAIN 1887 1976 IG-LIKE C2-TYPE DOMAIN 11.
 FT DOMAIN 1977 2070 IG-LIKE C2-TYPE DOMAIN 12.
 FT DOMAIN 2078 2162 IG-LIKE C2-TYPE DOMAIN 13.
 FT DOMAIN 2174 2260 IG-LIKE C2-TYPE DOMAIN 14.
 FT DOMAIN 2267 2347 IG-LIKE C2-TYPE DOMAIN 15.
 FT DOMAIN 2354 2436 IG-LIKE C2-TYPE DOMAIN 16.
 FT DISULFID 66 114 BY SIMILARITY.
 FT DISULFID 149 161 BY SIMILARITY.
 FT DISULFID 156 174 BY SIMILARITY.
 FT DISULFID 168 183 BY SIMILARITY.
 FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 197 215 BY SIMILARITY.
 FT DISULFID 209 224 BY SIMILARITY.
 FT DISULFID 223 246 BY SIMILARITY.
 FT DISULFID 240 259 BY SIMILARITY.
 FT DISULFID 253 268 BY SIMILARITY.
 FT DISULFID 954 963 BY SIMILARITY.
 FT DISULFID 956 970 BY SIMILARITY.
 FT DISULFID 973 982 BY SIMILARITY.
 FT DISULFID 985 1001 BY SIMILARITY.
 FT DISULFID 1010 1020 BY SIMILARITY.
 FT DISULFID 1012 1026 BY SIMILARITY.
 FT DISULFID 1029 1038 BY SIMILARITY.
 FT DISULFID 1041 1057 BY SIMILARITY.
 FT DISULFID 1060 1068 BY SIMILARITY.
 FT DISULFID 1062 1078 BY SIMILARITY.
 FT DISULFID 1081 1090 BY SIMILARITY.
 FT DISULFID 1093 1108 BY SIMILARITY.
 FT DISULFID 1151 1199 BY SIMILARITY.
 FT DISULFID 1337 1383 BY SIMILARITY.
 FT DISULFID 1434 1480 BY SIMILARITY.
 FT DISULFID 1526 1572 BY SIMILARITY.
 FT DISULFID 1617 1662 BY SIMILARITY.
 FT DISULFID 1718 1766 BY SIMILARITY.
 FT DISULFID 1813 1860 BY SIMILARITY.
 FT DISULFID 1906 1953 BY SIMILARITY.
 FT DISULFID 1997 2052 BY SIMILARITY.
 FT DISULFID 2098 2146 BY SIMILARITY.
 FT DISULFID 2194 2241 BY SIMILARITY.
 FT DISULFID 2283 2328 BY SIMILARITY.
 FT DISULFID 2373 2419 BY SIMILARITY.
 FT CARBOHYD 1421 1421 N-LINKED (GLCNAG).
 SQ SEQUENCE 2481 AA: 270812 MW: E47FBE2B310F5031 CRC64: (POTENTIAL).

Query Match 85.7% Score 30: DB 1: Length 2481;
 Best Local Similarity 83.3% Pred. No. 4.2e+02;
 Matches 5: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 PTPPG 6
 DB 711 PTPPG 716

RESULT 11
 LRP2_RAT STANDARD; PRU: 4660 AA.
 ID LRP2_RAT
 AC P98158;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 2 PRECURSOR (MEGALIN)
 DE (GLYCOPROTEIN 330).
 GN LRP2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RC [1]
 PC SEQUENCE FROM N.A.
 RP STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY;
 RX MEDLINE: 95024033.
 RA Saito A., Pletromonaco S., Ito A.K.C., Farquhar M.G.;
 RT Complete cloning and sequencing of rat gp330/megalin, a
 RT distinctive member of the low density lipoprotein receptor gene
 RT family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9725-9729(1994).
 RN [2]
 RP FUNCTION.
 RX MEDLINE: 95386696.
 RA Moestrup S.K., Cui S., Vorum H., Bregengaard C., Bjorn S.E.,
 RT Evidence that epithelial glycoprotein 330/megalin mediates uptake of
 RT polybasic drugs.";
 RL J. Clin. Invest. 96:1404-1413(1995).
 RN [3]
 RP TISSUE SPECIFICITY.
 RX MEDLINE: 94172242.
 RA Zheng G., Bachinsky D.R., Stamenkovic I., Strickland D.K., Brown D.,
 RT Andrus G., McCluskey R.T.;
 RT Organ distribution in rats of two members of the low-density
 RT lipoprotein receptor gene family, gp330 and LRP/alpha 2M, and the
 RT receptor-associated protein (RAP)."
 RL J. Histochem. Cytochem. 42:531-542(1994).
 CC -1- FUNCTION: BINDS PLASMINOGEN, EXTRACELLULAR MATRIX COMPONENTS,
 CC PLASMINOGEN ACTIVATOR-PLASMINOGEN ACTIVATOR INHIBITOR TYPE I
 CC COMPLEX, APOLIPOPROTEIN E-ENRICHED BETA-VLDL, LIPOPROTEIN LIPASE,
 CC LACTOFERRIN, CLUSTERIN AND CALCIUM.
 CC -1- FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS
 CC APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B.
 CC -1- SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR-
 CC ASSOCIATED PROTEIN (RAP).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXPRESSED IN
 CC CLATHRIN-COATED PITS; A SOLUBLE FORM IS POSSIBLY DERIVED BY
 CC CLEAVAGE AT THE CELL SURFACE.
 CC -1- TISSUE SPECIFICITY: EPITHELIAL CELLS OF KIDNEY GLOMERULUS AND
 CC PROXIMAL TUBULE, LUNG, EPIDIDYMIS, YOLK SAC, AMONG OTHERS.
 CC -1- SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 SH2-BINDING DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH2-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 CC EMBL: L34049; AAA51369.1; -.
 CC HSSP: P01130; IAIJ.
 DR INTERPRO: IPR000033; -.
 DR INTERPRO: IPR000152; -.
 DR INTERPRO: IPR000561; -.
 DR INTERPRO: IPR001881; -.
 DR INTERPRO: IPR002172; -.
 DR PFAM: PF00057; ldl_recept_a; 36.
 DR PFAM: PF00058; ldl_recept_b; 33.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PROSITE: PS00010; ASX_HYDROXYL; 4.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 8.

DR PROSITE: PS01187; EGF_CA: 3.
 DR PROSITE: PS01209; LDLRA_1; 31.
 DR PROSITE: PS50068; LDLRA_2; 36.
 KW Glycoprotein; Repeat; Endocytosis; Coated pits; Transmembrane;
 Receptor; EGF-like domain; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 4660
 FT DOMAIN 26 4425
 FT TRANSMEM 4426 4446
 FT DOMAIN 4447 4660
 FT DOMAIN 26 64
 FT DOMAIN 106 105
 FT DOMAIN 144 144
 FT DOMAIN 145 181
 FT DOMAIN 181 219
 FT DOMAIN 220 258
 FT DOMAIN 263 307
 FT DOMAIN 308 346
 FT DOMAIN 347 385
 FT DOMAIN 435 476
 FT DOMAIN 478 519
 FT DOMAIN 521 566
 FT DOMAIN 568 611
 FT DOMAIN 612 652
 FT DOMAIN 656 704
 FT DOMAIN 732 793
 FT DOMAIN 795 835
 FT DOMAIN 837 879
 FT DOMAIN 881 923
 FT DOMAIN 969 1013
 FT DOMAIN 1023 1061
 FT DOMAIN 1064 1103
 FT DOMAIN 1108 1146
 FT DOMAIN 1148 1186
 FT DOMAIN 1186 1225
 FT DOMAIN 1225 1269
 FT DOMAIN 1270 1308
 FT DOMAIN 1311 1351
 FT DOMAIN 1350 1390
 FT DOMAIN 1391 1430
 FT DOMAIN 1479 1520
 FT DOMAIN 1522 1563
 FT DOMAIN 1567 1609
 FT DOMAIN 1611 1654
 FT DOMAIN 1656 1696
 FT DOMAIN 1701 1742
 FT DOMAIN 1791 1832
 FT DOMAIN 1834 1882
 FT DOMAIN 1884 1930
 FT DOMAIN 1932 1972
 FT DOMAIN 1973 2013
 FT DOMAIN 2019 2060
 FT DOMAIN 2108 2156
 FT DOMAIN 2158 2201
 FT DOMAIN 2203 2245
 FT DOMAIN 2247 2289
 FT DOMAIN 2291 2332
 FT DOMAIN 2343 2384
 FT DOMAIN 2432 2477
 FT DOMAIN 2479 2518
 FT DOMAIN 2520 2562
 FT DOMAIN 2564 2604
 FT DOMAIN 2605 2647
 FT DOMAIN 2652 2694
 FT DOMAIN 2699 2739
 FT DOMAIN 2740 2778
 FT DOMAIN 2779 2820
 FT DOMAIN 2821 2862
 FT DOMAIN 2863 2903
 FT DOMAIN 2906 2947
 FT DOMAIN 2948 2992
 FT DOMAIN 2993 3031

POTENTIAL.
 LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED
 PROTEIN 2.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 LDL-RECEPTOR CLASS A 1.
 LDL-RECEPTOR CLASS A 2.
 LDL-RECEPTOR CLASS A 3.
 LDL-RECEPTOR CLASS A 4.
 LDL-RECEPTOR CLASS A 5.
 LDL-RECEPTOR CLASS A 6.
 LDL-RECEPTOR CLASS A 7.
 EGF-LIKE 1.
 EGF-LIKE 2.
 LDL-RECEPTOR CLASS B 1.
 LDL-RECEPTOR CLASS B 2.
 LDL-RECEPTOR CLASS B 3.
 LDL-RECEPTOR CLASS B 4.
 LDL-RECEPTOR CLASS B 5.
 EGF-LIKE 3.
 LDL-RECEPTOR CLASS B 6.
 LDL-RECEPTOR CLASS B 7.
 LDL-RECEPTOR CLASS B 8.
 LDL-RECEPTOR CLASS B 9.
 EGF-LIKE 4.
 LDL-RECEPTOR CLASS A 8.
 LDL-RECEPTOR CLASS A 9.
 LDL-RECEPTOR CLASS A 10.
 LDL-RECEPTOR CLASS A 11.
 LDL-RECEPTOR CLASS A 12.
 LDL-RECEPTOR CLASS A 13.
 LDL-RECEPTOR CLASS A 14.
 LDL-RECEPTOR CLASS A 15.
 EGF-LIKE 5.
 CALCIUM-BINDING (POTENTIAL).
 LDL-RECEPTOR CLASS B 10.
 LDL-RECEPTOR CLASS B 11.
 LDL-RECEPTOR CLASS B 12.
 LDL-RECEPTOR CLASS B 13.
 LDL-RECEPTOR CLASS B 14.
 EGF-LIKE 7.
 LDL-RECEPTOR CLASS B 15.
 LDL-RECEPTOR CLASS B 16.
 LDL-RECEPTOR CLASS B 17.
 LDL-RECEPTOR CLASS B 18.
 LDL-RECEPTOR CLASS B 19.
 EGF-LIKE 8.
 LDL-RECEPTOR CLASS B 20.
 LDL-RECEPTOR CLASS B 21.
 LDL-RECEPTOR CLASS B 22.
 LDL-RECEPTOR CLASS B 23.
 LDL-RECEPTOR CLASS B 24.
 EGF-LIKE 9.
 LDL-RECEPTOR CLASS B 25.
 LDL-RECEPTOR CLASS B 26.
 LDL-RECEPTOR CLASS B 27.
 LDL-RECEPTOR CLASS B 28.
 LDL-RECEPTOR CLASS B 29.
 EGF-LIKE 10.
 LDL-RECEPTOR CLASS A 16.
 LDL-RECEPTOR CLASS A 17.
 LDL-RECEPTOR CLASS A 18.
 LDL-RECEPTOR CLASS A 19.
 LDL-RECEPTOR CLASS A 20.
 LDL-RECEPTOR CLASS A 21.
 LDL-RECEPTOR CLASS A 22.
 LDL-RECEPTOR CLASS A 23.

FT DOMAIN 3032 3072 LDL-RECEPTOR CLASS A 24.
 FT DOMAIN 3075 3112 LDL-RECEPTOR CLASS A 25.
 FT DOMAIN 3113 3153 EGF-LIKE 11.
 FT DOMAIN 3154 3194 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 3241 3282 LDL-RECEPTOR CLASS B 30.
 FT DOMAIN 3284 3333 LDL-RECEPTOR CLASS B 31.
 FT DOMAIN 3335 3377 LDL-RECEPTOR CLASS B 32.
 FT DOMAIN 3379 3420 LDL-RECEPTOR CLASS B 33.
 FT DOMAIN 3421 3461 LDL-RECEPTOR CLASS B 34.
 FT DOMAIN 3467 3511 EGF-LIKE 13.
 FT DOMAIN 3512 3552 LDL-RECEPTOR CLASS A 26.
 FT DOMAIN 3553 3593 LDL-RECEPTOR CLASS A 27.
 FT DOMAIN 3594 3634 LDL-RECEPTOR CLASS A 28.
 FT DOMAIN 3635 3675 LDL-RECEPTOR CLASS A 29.
 FT DOMAIN 3678 3718 LDL-RECEPTOR CLASS A 30.
 FT DOMAIN 3719 3758 LDL-RECEPTOR CLASS A 31.
 FT DOMAIN 3759 3797 LDL-RECEPTOR CLASS A 32.
 FT DOMAIN 3798 3836 LDL-RECEPTOR CLASS A 33.
 FT DOMAIN 3842 3882 LDL-RECEPTOR CLASS A 34.
 FT DOMAIN 3883 3924 LDL-RECEPTOR CLASS A 35.
 FT DOMAIN 3928 3966 LDL-RECEPTOR CLASS A 36.
 FT DOMAIN 3968 4008 EGF-LIKE 14.
 FT DOMAIN 4009 4050 LDL-RECEPTOR CLASS B 35.
 FT DOMAIN 4156 4197 LDL-RECEPTOR CLASS B 36.
 FT DOMAIN 4199 4241 LDL-RECEPTOR CLASS B 37.
 FT DOMAIN 4244 4284 EGF-LIKE 16.
 FT DOMAIN 4332 4370 EGF-LIKE 17.
 FT DOMAIN 4379 4413 SH3-BINDING (POTENTIAL).
 FT DOMAIN 4413 4460 SH3-BINDING (POTENTIAL).
 FT DOMAIN 4457 4463 SH2-BINDING (POTENTIAL).
 FT DOMAIN 4606 4609 SH3-BINDING (POTENTIAL).
 FT DOMAIN 4619 4625 SH3-BINDING (POTENTIAL).
 FT DOMAIN 4624 4630 SH3-BINDING (POTENTIAL).
 FT DOMAIN 1743 1745 CELL ATTACHMENT SITE (POTENTIAL).
 FT DOMAIN 4522 4527 ENDOCYTOSIS SIGNAL (POTENTIAL).
 FT DISULFID 4601 4606 ENDOCYTOSIS SIGNAL (POTENTIAL).
 FT DISULFID 28 40 BY SIMILARITY.
 FT DISULFID 35 53 BY SIMILARITY.
 FT DISULFID 47 62 BY SIMILARITY.
 FT DISULFID 67 80 BY SIMILARITY.
 FT DISULFID 74 93 BY SIMILARITY.
 FT DISULFID 87 103 BY SIMILARITY.
 FT DISULFID 108 120 BY SIMILARITY.
 FT DISULFID 115 133 BY SIMILARITY.
 FT DISULFID 127 142 BY SIMILARITY.
 FT DISULFID 147 157 BY SIMILARITY.
 FT DISULFID 152 170 BY SIMILARITY.
 FT DISULFID 164 179 BY SIMILARITY.
 FT DISULFID 183 195 BY SIMILARITY.
 FT DISULFID 190 208 BY SIMILARITY.

Query Match 85.7%; Score 30; DB 1; Length 4660.
 Best Local Similarity 83.3%; Pred. No. 8.1e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Db 2635 PTPSG 2640

RESULT 12
 MEAL USTMA
 ID MEAL USTMA STANDARD: PRT: 40 AA.
 AC P31962.
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE A1-SPECIFIC PHEROMONE (MATING FACTOR A1).
 GN MEAL.
 OS Ustilago maydis (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 RN [1]

```
CC -----
DR EMBL: M1145; AAA29523.1; -.
DM PIR: A25780; A25780.
KM Signal; Malaria.
FT SIGNAL 1 16 PROBABLE.
FT CHAIN 17 162 CIRCUMSPOROZITE-PROTEIN RELATED ANTIGEN.
SQ SEQUENCE 162 AA; 17349 MW; FB0C08250BC0B589 CNC64;

Query Match
Best Local Similarity 82.9%; Score 29; DB 1; Length 162;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

QY 1 PROPOG 6
   | |||||
Db 146 PROPOG 151

RESULT 14
EXPL_PLAFA
ID EXPL_PLAFA STANDARD; PRT; 162 AA.
AC P04926; P06718;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MALARIA PROTEIN EXP-1 PRECURSOR (EXPORTED ANTIGEN AG 5.1).
Exp-1.
OC Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
[1]
```

Qy	1	PTOPQG	6		
Db	146	PEQPGQ	151		
RESULT 14					
EXP1_PLAFA					
ID	EXP1_PLAFA	STANDARD;	PTT;	162	AA.
AC	P04926; P06718;				
DT	13-AUG-1987 (Rel. 05, Created)				
DT	01-JAN-1988 (Rel. 06, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	MALARIAL PROTEIN EXP-1 PRECURSOR (EXPORTED ANTIGEN AG 5.1).				
OS	Exp-1.				
GN	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE; 87218504.				
RT	Simmons D., Woollett G., Bergin-Cartwright M., Kay D., Scaife J.,				
RT	"A malarial protein exported into a new compartment within the host				
RT	erythrocyte.";				
RL	EMBO J. 6:485-491(1987).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE; 87218504.				

KR	-The gene for an exported antigen of the malaria parasite Plasmodium falciparum cloned and expressed in Escherichia coli.";
KT	Nucleic Acids Res. 13:369-379(1985).
CC	-1 MISCELLANEOUS: THIS PROTEIN IS ADDRESSED TO A NEW COMPARTMENT WITHIN THE CYTOPLASM OF THE INFECTED RED CELL. IT ALSO SURROUNDS THE PARASITE, PROBABLY IN THE PARASTOPOPHOROUS VACUOLE MEMBRANE.
CC	-1 MISCELLANEOUS: THIS ANTIGEN AND THE CIRCUMSPOROZOITE PROTEIN APPEAR TO HAVE A COMMON EPITOPE.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their name and the source of the modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X05074; CAA28735.1; -.
DR	EMBL; X01745; CAA25881.1; -.
DR	EMBL; A04553; CAA00366.1; -.
DR	PIR; A23052; YAZ051.
DR	PIR; A26769; A26769.
KW	Malaria; Sporozoite; Transmembrane; Signal.
FY	SIGNAL
FY	1 22
FY	CHAIN 23 162 MALARIA PROTEIN EXP-1.
FY	TRANSMEM 80 101 POTENTIAL.
FY	DOMAIN 120 137 EPITOPES (DEDUCED).
FY	VARIANT 136 136 P -> G (IN 5.1 NEGATIVE STRAINS).
FY	VARIANT 160 160 D -> T (IN STRAINS PALO ALTOII & 3D7).
FY	CONFLICT 10 10 V -> A (IN REF. 2).
SS	SEQUENCE 162 AA; 17450 MW; AF6B5744CE4A4A12 CRC64;

Query Match 82.9%; Score 29; DB 1; Length 162;
 Best Local Similarity 83.3%; Pred. No. 41;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PROPOG 6
 1 1111
 DB 146 PROPOG 151

RESULT 15

SSB_SERMA STANDARD; PRT; 175 AA.
 ID SSB_SERMA
 AC P25762;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SINGLE-STRAND BINDING PROTEIN (SSB) (HELIX-DESTABILIZING PROTEIN).
 CN SSB
 OS Serratia marcescens
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SR41;
 RX MEDLINE: 93252278.
 RA de Vries J., Wackernagel W.;
 RT "Cloning and sequencing of the Serratia marcescens gene encoding a
 single-stranded DNA-binding protein (SSB) and its promoter region.";
 RL Gene 127:39-45(1993).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE: 95010041.
 RA de Vries J., Genschel J., Urbanke C., Thole H., Wackernagel W.;
 RT "The single-stranded-DNA-binding proteins (SSB) of Proteus mirabilis
 and Serratia marcescens.";
 RL Eur. J. Biochem. 224:613-622(1994).
 CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR REPLICATION OF THE
 CHROMOSOMES AND ITS SINGLE-STRANDED DNA PHAGES. IT IS ALSO
 INVOLVED IN DNA RECOMBINATION AND REPAIR.
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SSB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X65080; CAA46208.1; -;
 DR PIR: J06632; J06632.
 DR HSSP: P02339; 1KAW.
 DR INTERPRO: IPR000424; -;
 DR Pfam: PF00436; SSB_1.
 DR PROSITE: PS00735; SSB_1; 1.
 DR PROSITE: PS00736; SSB_2; 1.
 KW DNA-binding; DNA repair; DNA replication.
 FT INIT MET 0 0 BY SIMILARITY.
 FT DNA BIND 54 60 BY SIMILARITY.
 SQ SEQUENCE 175 AA; 18678 MW; 05F971ACE9B7239D CRC64;

Query Match 82.9%; Score 29; DB 1; Length 175;
 Best Local Similarity 83.3%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PROPOG 6
 1 1111
 DB 135 PROPOG 140

Search completed: November 8, 2000, 14:39:48
 Job time: 105 sec

Thu Nov 16 15:44:26 2000

us-09-308-237b-1_1.rsp

Page 11

RA Ha T.U., Pfeiffer F.J., Donahoe P.K.;
 RT "Human Ovarian Cancer, Cell Lines and Primary Ascites Cells, Express
 RT Clin. Cancer Res. 0:0-0(1999).
 DR EMBL: X91156; CAA62593.1; -.
 DR EMBL: X91157; CAA62593.1; JOINED.
 DR EMBL: X91158; CAA62593.1; JOINED.
 DR EMBL: X91159; CAA62593.1; JOINED.
 DR EMBL: X91160; CAA62593.1; JOINED.
 DR EMBL: X91161; CAA62593.1; JOINED.
 DR EMBL: X91162; CAA62593.1; JOINED.
 DR EMBL: X91163; CAA62593.1; JOINED.
 DR EMBL: X91164; CAA62593.1; JOINED.
 DR EMBL: X91165; CAA62593.1; JOINED.
 DR EMBL: X91166; CAA62593.1; JOINED.
 DR EMBL: X89013; CAA61418.1; -.
 DR EMBL: U28700; AAC50328.1; -.
 DR EMBL: AF172832; AAC68497.1; -.
 DR INTERPRO: IPR000472; -.
 DR INTERPRO: IPR000719; -.
 DR INTERPRO: IPR001245; -.
 DR PFAM: PF00069; PKINASE.1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 KW Signal; Receptor.
 FT SIGNAL 1 17 POTENTIAL.
 FT CDSIN 18 573 ANTI-MULLERIAN HORMONE TYPE II RECEPTOR.
 FT CONFLICT 161 L -> V (IN REF. 2).
 SQ SEQUENCE 573 AA: 62749 MW: 13470102942FDBA CRC64;

Query Match 100.0%; Score 35; DB 4; Length 573;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 PTPPOG 6
 Db 369 PTPPOG 374
 RESULT 2
 O9ZGN2 PRELIMINARY; PRT; 302 AA.
 AC O9ZGN2;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE OUTER MEMBRANE PROTEIN VAE1 (FRAGMENT).
 GN VAE1.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN: LGV-434;
 RA Bannantine J.P., Rocky D.D.;
 RT "Use of a primate model system for identification of Chlamydia
 RT trachomatis proteins recognized uniquely in the context of
 RT infection".
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBD databases.
 DR EMBL: AF077009; AAC35945.1; -.
 DR NON-REF 1 1
 FT NON-REF 1 1
 SQ SEQUENCE 302 AA: 33548 MW: 80297855C17ED6B1 CRC64;

Query Match 91.4%; Score 32; DB 2; Length 302;
 Best Local Similarity 83.3%; Pred. No. 38;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 PTPPOG 6
 Db 202 PTPPOG 207

RESULT 3
 O9V746 PRELIMINARY; PRT; 601 AA.
 ID O9V746
 AC O9V746;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE MSPO PROTEIN.
 GN MSPO.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyraoidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RX MEDLINE: 20196006.
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Gentler A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., Milshina N.V., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palzner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitzkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster".
 RL Science 287:2185-2195(2000).
 DR EMBL: AB003813; AAF58219.1; -.
 DR FLYBASE: FBgn020269; mspos.
 DR INTERPRO: IPR000884; -.
 DR INTERPRO: IPR002106; -.
 DR PFAM: PF000090; tsp.1; 1.
 DR PROSITE: PS00339; AA-TRNA_LIGASE-II-2; UNKNOWN.1.
 SQ SEQUENCE 601 AA: 65434 MW: 529EDDCD348AC8EF CRC64;

Query Match 91.4%; Score 32; DB 5; Length 601;
 Best Local Similarity 83.3%; Pred. No. 74;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 PTPPOG 6
 Db 338 PTPPOG 343

```

RESULT 4
ID 084244 PRELIMINARY; PRT; 792 AA.
AC 084244;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE OMP85 ANALOG.
GN YAPF.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE: 99000809.
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Mairathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL: AE001297; AAC67834.1; -.
DR INTERPRO: IPR000184; -.
DR PFAM: PF01103; Bac_surface_Ag; 1.
SQ SEQUENCE 792 AA; 88751 MW; 90916A8E43E9FD80 CRC64;

Query Match 91.4%; Score 32; DB 2; Length 792;
Best Local Similarity 83.3%; Pred. NO. 96;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPOG 6
Db 692 PTOPOG 697

RESULT 5
ID 09YH8 PRELIMINARY; PRT; 300 AA.
AC 09YH8;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HOMEDOMAIN PROTEIN.
GN MNR2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 98449697.
RA Tanabe Y., William C., Jessell T.M.;
RT "Specification of motor neuron identity by the MNR2 homeodomain protein."
RL Cell 95:67-80(1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AF066860; AAC64924.1; -.
DR HSSP: P14653; 1b2.
DR INTERPRO: IPR001366; -.
DR PFAM: PF00046; homeobox; 1.
DR PRINTS: PR00024; HOMEBOX.1.
DR PROSITE: PS00027; HOMEBOX.1; 1.
DR PROSITE: PS00711; HOMEBOX.2; 1.
KW Homeobox; DNA-binding; Nuclear protein.
SQ SEQUENCE 300 AA; 32300 MW; 19F710C312EC3E6 CRC64;

Query Match 86.6%; Score 31; DB 13; Length 300;
Best Local Similarity 83.3%; Pred. NO. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPOG 6

```

```

Db 238 PSOPG 243

RESULT 6
ID 062893 PRELIMINARY; PRT; 557 AA.
AC 062893;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE MIS TYPE II RECEPTOR.
GN RATMRIT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 96107132.
RA Teixeira J., He W.W., Shah P.C., Morikawa N., Lee M.M., Catlin E.A., Hudson P.L., Ming J., MacLaughlin D.T., Donahoe P.K.;
RT "Developmental expression of a candidate müllerian inhibiting substance type II receptor."
RL Endocrinology 137:160-165(1996).
DR EMBL: U42427; AAC52343.1; -.
DR INTERPRO: IPR000472; -.
DR INTERPRO: IPR00719; -.
DR INTERPRO: IPR001245; -.
DR PFAM: PF00069; Pkinase; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
SQ SEQUENCE 557 AA; 59748 MW; 8EDE9C0C32EBD05 CRC64;

Query Match 88.6%; Score 31; DB 11; Length 557;
Best Local Similarity 83.3%; Pred. NO. 11e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPOG 6
Db 367 PTOPOG 372

RESULT 7
ID 063045 PRELIMINARY; PRT; 557 AA.
AC 063045;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUT. ANTI-MÜLLERIAN HORMONE RECEPTOR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS.
RX MEDLINE: 94163972.
RA Baarends W.M., Van Helmond M.J.L., Post M., Van der Schoot P.J.C.M., Hoogerbrugge J.W., de Winter J.P., Gienbroek J.T.J., Karels B., Wilming L.G., Meijers J.H.C., Themmen A.P.N., Grootegoed A.J.;
RT "A novel member of the transmembrane serine/threonine kinase receptor family is specifically expressed in the gonads and in mesenchymal cells adjacent to the müllerian duct."
RL Development 120:189-197(1994).
DR EMBL: X71916; CAA50731.1; -.
DR INTERPRO: IPR00472; -.
DR INTERPRO: IPR00719; -.
DR INTERPRO: IPR001245; -.
DR PFAM: PF00069; Pkinase; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
KW Hormone.

```


QY 1 PTOPOG 6
 ||||:1
 Db 1124 PTOPOG 1129

RESULT 11

ID 008534 PRELIMINARY; PRT; 3074 AA.

AC 008534;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PUTATIVE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.44)
 DE (CONTAINS: 66 KDA PROTEIN).
 OS Sugar beet yellow virus (SBYV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
 OC Closterovirus.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-UKRAINIAN (BYV-U);
 RX MEDLINE: 94082464.
 RA Agronovsky A.A., Koonin E.V., Boyko V.P., Maiss E., Froetschl R.,
 Lunina N.A., Atabekov J.G.;
 RT "Beet yellows closterovirus: complete genome structure and
 identification of a leader papain-like thiol protease.";
 RL Virology 198:311-324(1994).
 CC -1- FUNCTION: PROBABLE POLYMERASE.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY A RIBOSOMAL FRAMESHIFT
 CC THAT OCCURS BETWEEN POSITIONS 2630 AND 2631.
 CC EMBL: X73476; CAA1871.1; -;
 DR INTERPRO: IPR000676; -;
 DR INTERPRO: IPR001788; -;
 DR INTERPRO: IPR002538; -;
 DR PFAM: PF00978; RNA_dep_RNApol2; 1.
 DR PFAM: PF01443; Viral_helicase1; 2.
 DR PFAM: PF01660; Vmethyltransf; 1.
 KM RNA-directed RNA polymerase.
 FT CHAIN 1 3074 348 KDA PROTEIN.
 FT CHAIN 1 3074 348 KDA PROTEIN.
 FT MUTAGEN 509 509 C->T: NOT PROCESSED INTO ITS SUBUNITS.
 FT MUTAGEN 517 517 C->T: REDUCED PROCESSING INTO SUBUNITS.
 FT MUTAGEN 518 518 C->T: REDUCED PROCESSING INTO SUBUNITS.
 FT MUTAGEN 556 556 H->E: NO EFFECT.
 FT MUTAGEN 559 559 H->E: NOT PROCESSED INTO ITS SUBUNITS.
 FT MUTAGEN 588 588 G->D: NOT PROCESSED INTO ITS SUBUNITS.
 SQ SEQUENCE 3074 AA; 346028 MW; D75B9564658BEB0C CRC64;

Query Match 88.6%; Score 31; DB 12; Length 3074;
 Best Local Similarity 83.3%; Pred. No. 6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPOG 6
 ||||:1
 Db 1124 PTOPOG 1129

RESULT 12
 ID 090703 PRELIMINARY; PRT; 3090 AA.
 AC 090703;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE P348.
 OS Beet yellows virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
 OC Closterovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BYV-4;
 RX MEDLINE: 20079557.
 RA Peremyslov V.V., Hagiwara Y., Dolja V.V.;

RT "HSP70 homolog functions in cell-to-cell movement of a plant virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776(1999).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-BYV-4;
 RA Peremyslov V.V., Hagiwara Y., Alzhanova D., Dolja V.V.;
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF190581; AAF14300.1; -;
 DR INTERPRO: IPR000606; -;
 DR INTERPRO: IPR001788; -;
 DR INTERPRO: IPR002588; -;
 DR PFAM: PF00978; RNA_dep_RNApol2; 1.
 DR PFAM: PF01443; Viral_helicase1; 1.
 DR PFAM: PF01660; Vmethyltransf; 1.
 SQ SEQUENCE 3090 AA; 347990 MW; 09C779A57DE874B9 CRC64;

Query Match 88.6%; Score 31; DB 12; Length 3090;
 Best Local Similarity 83.3%; Pred. No. 6e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTOPOG 6
 ||||:1
 Db 1124 PTOPOG 1129

RESULT 13
 ID 080529 PRELIMINARY; PRT; 213 AA.
 AC 080529;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE FLJ439.12 PROTEIN.
 GN FLJ439.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
 OC Brassicaceae; Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
 RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,
 RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,
 RA Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC003970; AAC33206.1; -;
 SQ SEQUENCE 213 AA; 22308 MW; 4216943C2CD3BBBF CRC64;

Query Match 85.7%; Score 30; DB 10; Length 213;
 Best Local Similarity 83.3%; Pred. No. 68;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTOPOG 6
 ||||:1
 Db 107 PTOPOG 112

RESULT 14
 ID 013015 PRELIMINARY; PRT; 297 AA.
 AC 013015;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE Y BOX PROTEIN 2.
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Cyprinidae; Cyprinidae; Cyprinidae; Cyprinidae;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE-OVARY;
 RX MEDLINE: 98055714.
 RA Katsun Y., Yamashita M., Nagahama Y.;
 KT "Isolation and characterization of goldfish y box protein, a germ-
 cell-specific RNA-binding protein.";
 RL Eur. J. Biochem. 249:854-861(1997).
 CC -1 SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
 DR EMBL: AB003336; BAA19850.1; -.
 DR HSSP: P32081; INMG.
 DR INTERPRO: IPR002059; -.
 DR PFAM: PF00313; CSD; 1.
 DR PRINTS: PRO0050; COLD_SHOCK.
 DR PROSITE: PS00352; COLD_SHOCK; 1.
 KW Transcription regulation; DNA-binding.
 SQ SEQUENCE 297 AA; 32775 MW; D006F95C8B76F33 CRC64;

Query Match 85.7%; Score 30; DB 13; Length 297;
 Best Local Similarity 83.3%; Pred. No. 95;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PPOPG 6
 1 1111
 DB 158 PPOPG 163

RESULT 15
 ID O9RV70 PRELIMINARY: PRT: 298 AA.
 AC O9RV70;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE URICASE.
 GN DR1160.
 OS Delinococcus radiodurans.
 OC Bacteria; Thermus/Delinococcus group; Delinococcales; Delinococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:RI;
 RX MEDLINE: 20036896.
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome Sequence of the Radioresistant Bacterium Delinococcus
 radiodurans RI.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001965; AAF10733.1; -.
 DR TIGR: DR1160; -.
 DR INTERPRO: IPR002042; -.
 DR PFAM: PF01014; URICase; 1.
 DR PRINTS: PRO0093; URICASE.
 SQ SEQUENCE 298 AA; 34332 MW; 1CCCT945F20F9C56 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 298;
 Best Local Similarity 83.3%; Pred. No. 95;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PPOPG 6
 1 1111
 DB 125 PPOPG 130